

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

# **How does the process of educational attainment differ between Africans and Coloureds in the Western Cape?**



**Clare Hofmeyr (HFMCLA001)**

This dissertation is presented in partial fulfillment of the requirements of the Masters of Commerce Degree in Applied Economics at the University of Cape Town, February 2011.

## ABSTRACT

More than a decade after the end of apartheid, inequality along racial lines is widely apparent. While the greatest disparity exists between Whites and non-Whites, inequality also persists among non-Whites. With reference to the youth of the Western Cape, Coloureds have higher per capita household income, more educated parents, superior schooling inputs and improved performance on achievement tests in comparison to Africans. The finding that Coloureds have only a slightly higher matriculation rate and a statistically equivalent enrolment rate leads one to ask whether the races face different influences on their process of educational attainment. Using individual and household data from the Cape Area Panel Study (CAPS), matriculation and enrolment (conditional on matriculation) are modeled for Africans and Coloureds separately. The role of current income and manifest ability are the focus of this analysis. Using logit regressions and predicted probabilities, the role of these two variables in educational attainment is examined. It is found that while manifest ability is important for both races at both stages, income has a differential impact by race. Coloureds display significant sensitivity to changes in current income while Africans do not. Looking at the entire range of income and ability reveals that Africans are between 20-40 percent more likely to matriculate than Coloureds when the races are set equal on all included characteristics. Discrimination in the labour market which favours Coloureds is suggested as a possible reason for this finding. At the tertiary level, however, enrolment by race does not differ significantly even when Africans and Coloureds are held equal, and is relatively low for both groups. While labour market factors are assumed to remain relevant, it is also suggested that African youths who grow up in poor areas of high unemployment may be unaware of the benefits which further education provides in the labour market.

## **ACKNOWLEDGEMENTS**

I would like to thank Professor Murray Leibbrandt and Associate Professor Cally Ardington for their invaluable support and advice during the writing of this dissertation. I would also like to thank Professor Leibbrandt and Professor David Lam for the opportunity to participate in the ICPSR summer courses at the University of Michigan, where I learnt several of the methods applied in this study.

I gratefully acknowledge funding from the National Research Foundation, the Mellon Foundation and Economic Research Southern Africa.

University of Cape Town

## PLAGIARISM DECLARATION

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.
2. This paper entitled "A comparison of African and Coloured educational attainment in the Western Cape" is my own work.
3. I have used the Harvard convention for citation and referencing. Each significant contribution to, and quotation in, this paper from the work(s) of other people has been attributed, and has been cited and referenced.
4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

Signature:

Date:

University of Cape Town

# TABLE OF CONTENTS

	Page
<b>1. Introduction</b>	<b>1</b>
<b>2. Literature Review</b>	<b>3</b>
2.1 Determinants of educational attainment	3
2.1.1 Family factors	3
2.1.2 Income	6
2.1.3 Ability	8
2.2 Education and the labour market in South Africa	10
2.2.1 Education in South Africa	10
2.2.2 Labour market factors	13
<b>3. Description of Data</b>	<b>17</b>
3.1 Attrition	18
3.2 Descriptive statistics	20
<b>4. Methodology</b>	<b>35</b>
4.1 Marginal effects	35
4.2 Predicted probabilities	37
4.2.1 Discrete changes in predicted probability	40
4.2.2 Predicted probability graphs	42
<b>5. Analysis</b>	<b>44</b>
5.1 Income	44
5.1.1 Marginal effects	44
5.1.2 Discrete changes in probability	52
5.1.3 Predicted probability graphs	56
5.2 Ability	63
5.1.1 Marginal effects	63
5.1.2 Discrete changes in probability	64
5.1.3 Predicted probability graphs	67
5.3 Interactions between income and ability	71
<b>6. Conclusion</b>	<b>74</b>
<b>7. References</b>	<b>76</b>
<b>8. Appendix</b>	<b>80</b>

# 1. INTRODUCTION

There is a large body of literature which documents the severity of racial inequality in South Africa, especially as it relates to education. It is commonly reported that the widest gaps exist between Whites and non-Whites. However, it is also true that inequality persists among non-Whites as well. In comparison to Africans, Coloureds hold the upper hand in almost every measure of socio-economic wellbeing: they have greater per capita household income, families are smaller with more educated parents, unemployment is substantially lower and the quality of schooling is higher (Lam et al 2007, Lam et al 2008). Given this, it is perhaps predictable that Coloured young adults (YAs) are found to fail fewer grades at school and score significantly higher on achievement test scores (Lam et al 2008).

What is not expected, however, is the relatively small advantage held by Coloureds in matriculation rates, and the lack of any significant difference in tertiary enrolment rates, conditional on passing Grade 12 (Bhorat and Oosthuizen 20008, Lam et al 2010). As educational attainment is commonly found to be strongly and positively correlated with household income, parental education and manifest ability (proxied by achievement test scores), why are Coloureds not outperforming Africans to a greater degree?

This study seeks to answer this question by looking at the process of educational attainment within the African and Coloured YA populations of the Western Cape using data from the Cape Area Panel Study (CAPS). This is a longitudinal survey with detailed information about the lives of youths as they leave school and enter the labour market. Given this detail, it is possible to control for the individual and household characteristics which are indicated in the literature to be relevant. Focus is placed specifically on the role of current household income and manifest ability.

The analysis uses several techniques related to logit regressions and predicted probabilities to answer three questions. Firstly, do ability and income play a significant role in the process of educational attainment once all other background factors have been accounted for? Secondly, how does the sensitivity to a change in income or ability differ by race? Thirdly, how does each race's responsiveness to income and ability affect their predicted probability of matriculation and enrolment over the entire range of these variables?

As racially unequal conditions in both education and the labour market are likely to play a role in the process of educational attainment but are not controlled for in the econometric analysis, the results are considered with them in mind. The nature and extent of inequality in education and the labour market are reviewed in Chapter 2, along with a discussion of the role which family factors, income

and ability are expected to play in educational attainment. Chapter 3 presents a descriptive analysis of the estimation sample, addressing attrition and providing a detailed look at how income and ability vary by race and level of education. This is followed by an outline of the methodology used in the analysis in Chapter 4, the results and discussion of which are presented in Chapter 5. Chapter 6 summarises the findings, suggests reasons for the racial patterns in educational attainment and suggests areas where government educational policy may be most effective for Africans and Coloureds.

University of Cape Town



## **2. LITERATURE REVIEW**

The first part of this chapter looks at how the process of educational attainment is conceptualised in the literature. Consideration will be given to all of the factors thought to be important, bearing in mind their relationship to income and ability, the variables of interest for this study. In the second section, the state of the South African education system and labour market will be presented. This will provide an idea of how one might expect the relative levels and responsiveness of educational attainment to vary by race given the realities of the schooling system and labour market in South Africa.

### **2.1 Determinants of educational attainment**

In order to accurately estimate the role of income and ability in matriculation and enrolment in post-secondary education (PSE), it is necessary to account for any other factors or processes which are likely to have an impact on educational attainment, and which could bias the estimates of income and ability if they are not taken into account. In addition, as many of the determinants of educational attainment are inextricably linked, it is necessary to understand these links so as to place the role of income and ability in context.

It should be noted that in the course of this literature review, papers which deal with the determinants of enrolment into PSE appeared to be more numerous than those which look at the completion of high school. Dustmann (2004) and Belley and Lochner (2007) have noted this phenomenon. While this may seem problematic given that this study looks at matriculation in addition to enrolment, it appears that by and large the same factors are hypothesised to determine both stages of educational attainment. This review will therefore discuss the determinants of educational attainment in general. Nonetheless, this is not true of South Africa and the empirical analysis of this paper works hard to deal with the fact that the attainment of matric and the enrolment into tertiary education are two different processes that, at least potentially, have different determinants.

#### **2.1.1 Family factors**

In the vast majority of the studies dealing with the determinants of young adults' educational outcomes, family background is given considerable attention, in particular parental characteristics (e.g. Belley and Lochner 2007, Bhorat and Oosthuizen 2008, Dustmann 2004, Fehrman et al 1987,

Jacoby 1994, Kodde and Ritzen 1988, Lam 1999, Lam et al 2008, Lauer 2003, and Mora 1997, among many others).

The parental characteristics most often cited relate to education and occupation. While the former appears in almost every study, the latter is accounted for only in the sociology literature. As occupation is a proxy for long-term income, and is generally highly correlated with parental education, the economics literature includes only education to capture the financial situation of the household in the long-run.

With regard to parental education, the majority of the literature finds a significant, positive relationship between the education of parents and their children (Albert 2000, Baumann 2007, Beattie 2002, Bratti 2002, Dustmann 2004, Hung et al 2000, Kodde 1987, Kodde & Ritzen 1988, Lauer 2003, Tomkowicz and Bushnik 2003). Dustmann (2004) suggests that children of parents with higher education may be more likely to complete high school and enrol in tertiary education because they have inherited greater learning ability from their parents or because their parents have directed them on that path due to their own preferences for higher education. Lam (1999) suggests that parents with more education may be better placed to help their children with their school work, they may indirectly pass on factors such as improved language skills, they may be able to afford better quality schooling due to holding more highly paid employment and there may be a neighbourhood effect due to the tendency of more educated parents to reside in areas with a higher average level of schooling. Disentangling these effects is difficult, if not impossible in some cases (Dustmann 2004). However, as the focus of this study is not on how parental factors influence educational attainment, this disentanglement is not relevant here. More pertinent is how the estimated effects of income and ability may be biased if they reflect aspects of omitted parental education.

As parental education is positively correlated with both income (as more educated parents hold better-paid jobs on average), and children's educational outcomes, omitting parental education could upwardly bias the estimated effect of income. In addition, families with more educated parents are expected to have higher income in the long-term, which is associated with improved household resources and the ability to afford better quality schooling (Lam 1999). As both of these factors are associated with improved educational outcomes, parental education is often included as a proxy for long-term family income in order to separate out its effects from current income (Bratti 2002, Carneiro and Heckman 2002, Dearden et al 2004, Lam et al 2010). This will be discussed in more detail in the following section. This study is interested largely in the effect of current income, as it is more open to government intervention (Bratti 2002). It will therefore be necessary to account

for parental education in this analysis in order to prevent an upward bias of the estimated effect of income.

Furthermore, parental education is linked to attitudes and preferences regarding educational attainment in several ways. More educated parents may prioritise the education of their children to a greater degree than otherwise identical but less educated parents (Anderson et al 2001, Dustmann 2004, Lauer 2003, Tomkiewicz and Bushnik 2003). This will affect both the decisions taken by parents regarding their child's educational path as well as the child's own attitude towards education (de Broucker 2005), and on average is expected to lead to higher levels of educational attainment. To the extent that parental education is determined by the parents' cognitive ability, and to the extent that this ability is passed on to children, omitting parental education may cause the effect of parental attitude to be reflected in the measure of the child's ability. It may also be picked up by income due to the aforementioned positive correlation between parental education and family income. In either case, the estimated effect of income or ability is likely to be biased if parental education is omitted from the analysis.

Aside from characteristics relating directly to parents, it is also important to account for family structure. This refers to the presence of parents in the household as well as the number of siblings (Anderson 2000, Astone and McLanahan 1991, Downey 1995). With regards to the latter, the probability of educational attainment at both the secondary and tertiary level has been found to decline with the number of siblings, even when parental education, socio-economic status and family income are controlled for (Alexander et al 1997, Tomkiewicz and Bushnik 2003).

Downey (1995) explains this relationship with a theory termed resource dilution, which posits that parents have a finite supply of resources to devote to child care, such as income, time and energy. These resources are associated with better educational outcomes, and the more children a family has, the thinner these resources have to be spread. Downey mentions other characteristics that are correlated with family size, such as socio-economic status and parental education; individuals from higher social classes who are relatively more educated tend to have fewer children. Thus the omission of family size could bias variables such as parental education and more importantly, for this study, family income and ability.

Turning to parental co-residence, Anderson (2000) shows that in comparison to youth living with both parents, youth living in all other home situations are less likely to be enrolled in school, and if enrolled, have completed fewer grades on average. Astone and McLanahan (1991) find that children living with single parents or step families receive less parental encouragement and attention than

children in intact families, and that family structure affects progress through school<sup>1</sup>. Finally, Fehrman et al (1987) find that parental involvement, which incorporates many different types of activities such as verbal encouragement, expectations and helping with school work, has both a direct and indirect effect on performance in the last year of high school. Clearly parental involvement is likely to be significantly limited if one or both parents are not co-resident in the child's household.

Failing to account for the presence of parents in the household could result in the income measure picking up the effect of family structure and biasing it upwards. For example, Astone and McLanahan (1991) suggest that non-intact families suffer from lower financial security. Children from 'broken' homes would therefore perform more poorly at school because of both limited parental involvement and limited family income.

The preceding discussion makes it clear that there are numerous ways in which parental characteristics and family structure affect educational outcomes. As such, all of the estimation work in this study accounts for these factors in order to estimate the effects of income and ability on educational attainment as accurately as possible.

### **2.1.2 Income**

There are a number of studies which attempt to explain the relationship between family income and educational attainment (e.g. Belley and Lochner 2007, Bratti 2002, Carneiro and Heckman 2002, Dearden et al 2004, Jacoby 1994 and Lam et al 2010). Belley and Lochner (2007) provide a brief explanation of two theories. The first posits the existence of a consumption value of education, implying that better-off individuals will be willing and able to pay more in order to acquire more education. The second theory suggests that those in higher socio-economic strata are more likely to obtain more education because their peers hold more education, their social networks allow them access to jobs which require more education and they are more aware of the relative benefits of further education.

Neither of these alternative theories receives much attention in the literature. Instead, many authors focus on separating out long-term and short-term income in order to ascertain the prevalence of short-term credit constraints. Carneiro and Heckman (2002) use data from the late 1970s to test for credit constraints in the United States. A number of authors follow their

---

<sup>1</sup> The authors note that the link between family encouragement and progress through school is not found to be significant and that the relationship between family structure and educational outcomes is therefore due to another, as yet unknown, channel.

methodology to determine the presence of short-term credit constraints in different countries, and at different points in time (Belley and Lochner 2007, Bratti 2002, Dearden et al 2004, Lam et al 2010). The following explanation of this theory draws on the work of all of the aforementioned authors.

According to the theory of credit constraints, the positive correlation between family income and educational attainment can be interpreted in two ways. The first interpretation considers the effect of long-term credit constraints on educational attainment: children who come from relatively low-income families are likely to have poor household resources and attend lower quality schools. Over the years, this long-term deprivation negatively affects the acquisition of cognitive and non-cognitive skills which play an important role in educational attainment. Thus the observed positive relationship between family income and educational attainment is due to the long-term effects of income on the development of the child. The second interpretation is simpler and more open to policy intervention. The relationship between income and education is explained in terms of short-term credit constraints: families which have less income are more constrained in their ability to finance education for their children. Thus the relationship between income and education derives from the fact that higher-income families are able to afford further education for their children while poorer families are not.

Some of the studies which use data from the U.S. and U.K. consider the role of credit constraints at the post-secondary stage alone (Bratti 2002, Carneiro and Heckman 2002). This may be because free secondary school education is widely available in these countries, while tertiary education is relatively expensive. In other words, credit constraints (at least in the short-term) are not expected to be relevant to secondary schooling. However, this study will examine the importance of family income at both the secondary and post-secondary level. This is in line with Belley and Lochner (2007) and Dearden et al (2004), and potentially provides a link between the findings of Lam et al (2008), who look at progress through secondary school, and Lam et al (2010), who deal with the importance of credit constraints at the post-secondary level.

The method suggested by Carneiro and Heckman (2002), and employed by the aforementioned authors, involves isolating the effect of short-term income by including variables in the estimation which capture ability, parental education, and family structure. It is argued that parental education is a strong predictor of long-term income, which in turn determines the family and schooling environment in which a child grows up. In Carneiro and Heckman's own words, long-run family and environmental factors are "crystallized" in the ability of the child (2002: 707), and by accounting for

both ability and parental background, one is able to isolate the relationship between current income and educational attainment.

While the findings regarding the role of credit constraints vary by context (these will be discussed in the section covering empirical findings), suffice it to say here that it appears necessary to account for long-term factors, such as family background and ability, in order to isolate the role of current income in the process of educational attainment. This is the approach taken in the analysis of Chapter 5.

### **2.1.5 Ability**

The concept of ability has been mentioned numerous times already but is yet to be clearly defined in the context of this study. It is not an easy concept to pin down, due to the debate surrounding innate versus manifest/measured ability. Hansen et al (2004) tackle this issue directly by determining whether or not schooling affects achievement test scores (which are often used as a measure of ability). They present two mutually exclusive views of the relationship between ability and schooling. The first claims that cognitive ability is fixed from around the age of 8, after which additional schooling is assumed to have no effect. This implies that IQ tests and achievement tests (such as the literacy and numeracy evaluation that will be used in this study) which are administered after this age measure the same thing. The authors cite Herrnstein and Murray (1994) as proponents of this view.

The second view suggests quite a different relationship between schooling and measured ability. Here it is assumed that schooling cannot affect latent ability, because it is genetically determined and set at birth, but that it can affect measured ability. In other words, schooling will have no effect on IQ but will impact scores attained on achievement tests throughout primary and secondary schooling. This implies that better quality schooling will have a greater effect on achievement test scores than lower quality schooling, all else being equal. Neal and Johnson (1996) are cited as an example of a paper which supports this second view.

In their analysis, Hansen et al (2004) conclude that schooling does indeed affect achievement test scores. They conceptualise the schooling-ability relationship as follows: measured ability is determined by both latent ability and schooling, with the latter also being determined by latent ability. Their econometric analysis indicates that an additional year of schooling has a roughly equal effect on achievement test scores throughout primary and secondary schooling. They also find that schooling has the greatest effect for those with the lowest innate ability, indicating that it can act as an equaliser to some extent.

Hansen et al (2004) warn that ignoring the bi-directional causality between ability and schooling may lead to overestimating the role of innate ability in predicting educational attainment. However, this study is interested not in innate ability, which cannot be altered, but rather in measured ability which appears to be more open to government intervention. By comparing the importance of income and measured ability in the educational attainment of Africans and Coloureds, light may be shed on how this process differs by race. Furthermore, it may be possible to highlight areas where government policy might be most effective.

The role of family background in determining measured ability is much less contested than that of schooling, and has already been referred to several times in the preceding discussion. Dustmann (2004) and Lam (2009) argue that children may acquire improved language skills and a greater ability to learn from more highly educated parents. Carneiro and Heckman (2002) provide a similar argument, claiming that ability partly reflects long-run family and environmental factors. Finally, Heckman et al (2006) cite a number of papers which find that the cognitive ability of a child is strongly influenced by both the ability and involvement of their parents. It is therefore assumed in this paper that measured ability is affected by both family background and the quality and quantity of schooling received.

Thus far the discussion has dealt with cognitive ability, as achievement tests measure skills related to literacy and numeracy. Heckman et al (2006) suggest non-cognitive ability may also play a role in educational outcomes. Non-cognitive ability refers to characteristics such as persistence, motivation and charm. A number of studies from the fields of sociology, psychology and even Marxist economics find that these skills are important determinants not only of educational outcomes, but also of occupational attainment and wages (Heckman et al 2006).

In their analysis, Heckman et al (2006) use scores from two tests to capture non-cognitive ability: the Rotter Locus of Control Scale, and the Rosenberg Self-Esteem Scale. They acknowledge that these tests may be imperfect measures of non-cognitive ability, but they are limited by the available data. It is unfortunate that no such measures are available in the CAPS dataset used here. The only available test score measures literacy and numeracy abilities, as mentioned previously. This means that it is impossible to test the relative importance of cognitive and non-cognitive ability in educational attainment.

To the extent that the two types of abilities are correlated, the failure to control for non-cognitive skills may bias the effect of cognitive ability. Furthermore, Heckman et al (2006) argue that the omission of non-cognitive skills may upwardly bias the estimated effects of family background as

many of these skills are learnt at home. However, this second bias is not of relevance, as parental factors are incorporated simply as controls. Unfortunately nothing can be done with regards to the possible bias in the effect of cognitive ability, except to acknowledge that it may exist. As Heckman et al (2006) conclude that both non-cognitive and cognitive ability are determined by many of the same factors (parenting and schooling in particular), and that both are predictors of educational attainment, it does not seem overly problematic that the included measure of cognitive ability may pick up effects non-cognitive ability. In terms of policy interventions, improved schooling will raise both types of skills and thereby raise educational attainment.

In summary, the estimations in Chapter 5 will look at the role of measured rather than innate ability in educational attainment, where it is assumed that measured ability is determined by schooling and family factors. This suits the more practical orientation of the analysis, which seeks to understand the mechanics of educational decision-making among Africans and Coloureds, as well as pointing to the sorts of government policies which may be most effective in improving educational outcomes.

## **2.2 Education and the labour market in South Africa**

The second section of this chapter considers the empirical findings regarding the state of education and the labour market in South Africa which may have a bearing on the relationship between income, ability and educational attainment. This will place the process of educational attainment in the appropriate context, and will highlight the factors which are expected to play a role in schooling decisions and outcomes.

### **2.2.1 Education in South Africa**

As many of the problems still challenging youth and policy makers of today are hangovers from apartheid, this section begins with a brief analysis of the education system pre-1994. As Bray et al (2010) explain, schooling under apartheid was subject to deep inequalities on almost every basis; Whites were educated in the manner of a first world country while non-Whites, Africans in particular, received schooling comparable to some of the poorest African nations. For example, schooling was not compulsory for Africans, the quality of teaching was extremely poor and often relied on rote-learning, enrolment levels were low and drop-out rates were high. In 1970, secondary school enrolment stood at just 16 percent for Africans as opposed to 90 percent for Whites, with Coloured students falling somewhere in-between (Bray et al 2010: 170). While Bray et al (2010) state



that the racial gap in enrolment rates and real expenditure began to decline two decades before the end of apartheid, it is also true that as recently as the late 1980s, educational expenditure per White child was four times greater than per Indian or Coloured child, and ten times greater than for an African child (Moll 1998: 263).

Strides have certainly been made towards equal schooling for all, most notably in terms of government expenditure. By 1997, for every R1 spent on a White school pupil, R1.41, R1.27 and R1.39 was spent on each Indian, Coloured and African school pupil respectively (van der Berg 2006: 635). Furthermore, the democratic government appears to have met almost all of its specified targets: schooling is no longer racially segregated, enrolment until Grade 9 was made compulsory for all in 1996, new schools have been built, a new curriculum was introduced in 1997 (and then again in 2004) and numerous investments have been made in teacher training (Bray et al 2010).

Despite these changes, racial disparities in educational inputs remain. Using data from 2000, Bhorat and Oosthuizen (2008) find that Coloureds fared better than Africans on most measures of schooling inputs. For example, they find that a greater proportion of Coloured schools charged fees (2008: 640), allowing more teachers (or possibly better quality teachers) to be hired. This is reflected by the fact that the ratio of private teachers<sup>2</sup> to public teachers was substantially higher for Coloured schools than African schools. Coloured schools also fare better in terms of resources, such as specialist classrooms, libraries and computers.

In addition to unequal resources, several authors conclude that the quality of management and teaching in schools still varies considerably by race (Bhorat and Oosthuizen 2008, Crouch and Mabogoane 2001). Unfortunately these factors are difficult to quantify. Bhorat and Oosthuizen (2008) find that the presence of part- or fully-funded staff accommodation has a positive influence on Grade 12 pass rates. They use this to suggest that staff accommodation may be used as a proxy for the quality of teaching, as it raises the real wage and therefore may attract higher quality teachers. Crouch and Mabogoane (2001) do not suggest a proxy, but rather argue that teacher and management quality must explain the difference in outcomes which is evident even when resources are held constant.

Whether educational outcomes are driven more by material resources or by the quality of teaching and management is as yet unclear. However, it is uncontroversial that educational inputs remain racially unequal well past the end of apartheid. The fact that the majority of each race still attends the same schools they would have been forced to pre-1994 (Yamauchi 2005), and that a school's

---

<sup>2</sup> Private teachers are also referred to as governing-body teachers, as they are hired by the governing body of the school rather than the government.

former department under apartheid still plays a significant role in predicting Grade 12 pass rates (Bhorat and Oosthuizen 2008), are clear indicators of this.

As a consequence, educational outcomes also remain differentiated by race. Lam et al (2008a) find that Africans have a high rate of grade repetition, but are also very likely to be enrolled in school into their late teens or even 20s, and accordingly have a low dropout rate. Coloureds have a lower grade repetition rate but are much more likely to drop out. These findings alone do not indicate how the final educational attainment of Africans and Coloureds compare. For this, it is necessary to look at matriculation and enrolment rates by race.

Using data from the Department of Education, Bhorat and Oosthuizen find that 75 percent of Coloured pupils who wrote matric in 2000 passed, as compared to only 48 percent of African pupils (2008: 640). Lam et al (2010) use a sample of 25-29 year-olds from the Labour Force Survey to calculate the proportion of Africans and Coloureds with matric in each of the years between 2000 and 2007. They find that 28 percent of both groups held matric in 2000, rising to 33 percent for Africans and 38 percent for Coloureds in 2007.

While the findings of Bhorat and Oosthuizen (2008) might lead one to conclude that Coloureds substantially outperform Africans in terms of educational attainment, those of Lam et al (2008) indicate otherwise. This apparent contradiction is likely due to high levels of grade repetition among Africans. As African pupils remain in school for substantially longer than Coloureds, it is likely that they may attempt matric more than once, and may pass on their second or third attempt. This means that the inequality in grade completion at younger ages does not seem to translate into vastly different *final* matriculation rates.

While the government has used the increase in the matric pass rate since the end of apartheid as proof of improvements in the schooling system, Bray et al (2010) argue that this rise is due in part to fewer pupils taking the matric exams in the years between 1998 and the early 2000s. Crouch and Vinjevold (2006) lend support to this argument when they state that in an attempt to meet national pass rate targets set by the Minister of Education in 2001, provincial education departments discouraged students who were likely to fail from attempting the matric exams at all. The authors also mention that more students are taking subjects such as maths and science on standard grade rather than higher grade, or opting to take other subjects altogether. This means that the value of passing matric is devalued if it is attained by taking less cognitively-demanding material.

The low rate of matriculation, in combination with the declining standard of the qualification, has dire implications for the level of preparedness for tertiary education (Crouch and Vinjevold 2006).

This is reflected by statistics of enrolment in and completion of some form of post-secondary qualification. Lam et al (2010: 21) report that 7.4 and 7.5 percent of Africans aged 25-29 had completed some level of further education in the years 2000 and 2007 respectively, while the figures stood at 7.9 and 6.8 percent for Coloureds<sup>3</sup>. Using data from the Cape Town metropolitan area, the authors report that 30 percent of Africans and 27 percent of Coloureds who had matriculated between the years 2002-2004 went on to enroll in some form of tertiary education in 2005 (2010: 24). These two sets of statistics are not strictly comparable as the former refer to South Africa as a whole while the latter only to a section of the Western Cape. However, the discrepancy could also be explained, in part, by the fact that enrolment rates are likely to be higher than completion rates due to dropouts. In general, these results indicate that while inequality between Africans and Coloureds at the post-secondary level may be low or non-existent, the completion of tertiary studies remains woefully low for both groups.

The preceding discussion makes it clear that more than a decade after the end of apartheid, educational inputs remain unequal among non-Whites; Coloureds fare better than Africans in terms of school resources and quality of teaching and management. This would lead one to expect higher levels of educational attainment among Coloureds than Africans. Interestingly, the racial gap is relatively small in reality. The proportion of Coloured YAs with matric only slightly exceeds that of Africans in the first decade of the 21<sup>st</sup> century, while enrolment rates are equal, if not slightly higher for Africans. This is likely to be due to the fact that while Coloureds have a lower failure rate than Africans, they have a greater propensity to drop out. This study seeks to understand why this is the case. It is likely that the state of the labour market has a part to play, and it is to this that we now turn.

## **2.2.2 Labour market factors**

A number of papers control for labour market conditions when studying the process of educational attainment (Albert 2000, Gustman and Steinmeier 1981, Hansen et al 2004, Heckman et al 2006, Hung et al 2000, Kodde 1986, Kodde 1987, Lam et al 2008, Willis and Rosen 1979). Specifically, all of these analyses consider the role of actual or expected wages, as well as unemployment rates, in the process of completing secondary education and going on to enrol in PSE.

---

<sup>3</sup> In comparison, the proportion of Whites in this age group holding some form of tertiary education was four to five times greater in both years. Note that these findings are derived from the Labour Force Survey and hence are nationally representative.

The role of the labour market in educational attainment relates to the opportunity cost of time; the longer a YA stays in school, the longer they go without earning a wage or, at least, a full-time wage (Lam et al 2008). Thus there is an incentive to drop out of school in order to enter the labour market. However, Lam et al (2008) explain that a second effect may work in the opposite direction: if additional schooling raises the probability of employment or improves the type of employment likely to be found, YAs may choose to stay on and complete secondary schooling. This line of reasoning can be extended to tertiary education. Which effect dominates is necessarily an empirical question.

South Africa is a case in point of an economy which has vastly different labour market outcomes by education level, and a number of studies have focused explicitly on this issue. Bhorat and Leibbrandt (2001) find that completed secondary schooling significantly improved the probability of finding employment in 1995. Dias (2005) looks at the same issue in 2003 and finds that matric no longer provided any additional protection from unemployment over and above a primary school education. Only those with some form of tertiary education were estimated to be more likely to find employment<sup>4</sup>. Lam and Leibbrandt (2009) further update this work to 2007. Referring to Figure 3 of that paper, the authors show that the proportion of African men aged 25-29 finding work rose slowly as education increased from zero to 12 years. Once matric is attained, the probability of employment increased much more sharply with further education<sup>5</sup> (2009: 16).

Using a multinomial logit regression, Branson et al (2009) provide econometric support for Lam and Leibbrandt's (2009) findings. They consider the returns of schooling to finding wage- and self-employment for all years between 2000 and 2007 (inclusive). It is estimated that throughout this period, 25-29 year olds with tertiary education managed to find wage employment significantly more easily than matriculants, who in turn found wage employment more easily than non-matriculants. Holding a matric did not make a significant difference in terms of finding self-employment, whereas tertiary education did in all years except 2000 and 2003<sup>6</sup>. They also demonstrate that the return to tertiary education in terms of employment rose between 2000 and 2007. While those with tertiary education were twice as likely to be employed as those with matric

---

<sup>4</sup> Dias's results also indicate that labour force participants with no education are significantly more likely to find employment than those who hold a primary school education. She presents statistics regarding work experience which suggest that labour force participants who have no education are older and more experienced on average, and for this reason may be in greater demand in the labour market (2005: 22).

<sup>5</sup> Note that the results of Lam and Leibbrandt (2009) which are discussed in this and the following paragraphs are descriptive and therefore do not control for any other factors.

<sup>6</sup> The fact that Branson et al (2009) distinguish between wage and self-employment while Dias (2005) does not might explain the discrepancy regarding the estimated effect of matric on the probability of finding employment.

in 2000, by 2007 this had risen to three times as likely. In contrast, the protection afforded by matric stayed more or less constant over this period.

Relative earnings by education level can also be expected to play a role in the process of educational attainment. Looking at African men between the ages of 25-29 in 2007, Lam and Leibbrandt (2009) find that the earnings of an employed matriculant are 29 percent higher on average than those of a worker with only a Grade 11 education. For those with matric, earnings increase at a much higher rate with further education. Africans holding two years of post-secondary education earned roughly 100 percent more than those with only a matric.

These findings are once again supported by the econometric analysis of Branson et al (2009). Using Ordinary Least Squares, the authors find that in relation to those with less than complete secondary education, earnings are between 40 and 70 percent higher for matriculants and between 170 and 400 percent higher for those with post-secondary education over the period 2000-2007 (2009: 12).

As this paper is interested in a comparison of educational attainment between Africans and Coloureds, it is necessary to consider employment experiences by race. Should Africans and Coloureds face different probabilities of finding work and different earnings once employed, the opportunity cost of staying on in full-time education is racially differentiated.

The literature indicates that Africans and Coloureds face markedly different labour market opportunities. Lam et al (2009) present figures of employment for Africans and Coloureds living in Cape Town and aged 19-20 in 2002. They find that between 8-12 percent of Africans were working in 2002 as opposed to 43-52 percent of Coloureds (the figures are disaggregated by gender). In terms of having worked at all in the last 12 months, the gap is even larger: 12-18 percent of Africans report having worked versus 58-73 percent of Coloureds.

Lam et al (2007) provide further evidence of the stark racial gap in labour market outcomes. Using data from Waves 1-4 of CAPS, the authors find that two years after having left school, the proportion of Africans ever having worked was approximately 50 percent, and that 40 percent were currently working. On the other hand, approximately 85 percent of Coloureds had ever worked and close to 70 percent were currently working. Unfortunately the racial gap also extends to wages. Among school-leavers aged 20 or older, average earnings of African men were just over half those of Coloured men. While the Africans in this sample lag behind slightly in terms of average schooling, the authors suggest that this is only part of the reason for the relatively large gap in earnings.

Lam et al (2007) suggest that the extreme spatial segregation created during apartheid may be one reason for the differential labour market outcomes. Africans tend to live far away from business centres, while Coloureds traditionally live much closer. It is therefore easier for Coloureds to find employment. Furthermore, it is likely that a preference for hiring Coloureds as a result of labour practices in the Western Cape during apartheid endures into the 21<sup>st</sup> century (Lam et al 2007).

It is clear that there are a number of aspects of the South African labour market which are likely to influence educational attainment. Those who hold more education face a greater likelihood of employment and can expect to earn higher wages once working. This is especially evident for individuals with some form of post-secondary qualification, and is true for both races. This might be expected to encourage matriculation and enrolment. However, on average Coloureds tend to find jobs much more quickly and easily than Africans, and are likely to earn more once working. This means that Coloureds face a much higher opportunity cost to staying in school and going on to enrol in PSE. All of these factors will be considered when estimating the relationship between educational attainment and income and ability.

### 3. DESCRIPTION OF DATA

The data used in this analysis come from the Cape Area Panel Study (CAPS), a longitudinal survey started in 2002 with five waves currently completed. The study tracks the lives of YAs as they leave school and begin working. The first wave included 4752 individuals between the ages of 14-22 living in the Cape Town metropolitan area (Lam et al 2010). One of the great advantages to this dataset is a level of detail which is missing from most other large surveys. Information is collected relating to demographics, household resources and expenditure, family characteristics, sexuality, education and work. This analysis utilises data from Waves 1, 3, 4 and 5 (which were captured in 2002, 2005, 2006 and 2009 respectively). The period of analysis is restricted to 2002 - 2006 with data from Wave 5 used to fill in the gaps for respondents who were not surveyed in either Wave 3 or 4.

Of particular relevance to this analysis is the literacy and numeracy evaluation (LNE) which was administered to all YAs in the first wave. As Lam et al (2009) explain, this was a relatively short test which could be taken in English or Afrikaans. The authors point out that the vast majority of Africans did not take the test in their mother tongue while the opposite is true of Coloureds. This may well have had a negative impact on the scores of Africans. However, given that the language of instruction in African schools is officially English, and that matric exams are taken in English (Lam et al 2009), writing the evaluation in this language should not have placed Africans at too great a disadvantage.

In this study, the LNE score is used as a measure of manifest or measured ability. This is in line with Lam et al (2009) and Lam et al (2010). The scores reflect innate ability as well as acquired knowledge and skills. The latter are assumed to be affected by the quantity and quality of schooling as well as the family background of the individual.

The full sample on which the Stage 1 or matriculation regressions are run is comprised of 1289 individuals. It includes all those who left school between 2002 and 2005 (inclusive). The YAs in this sample passed matric, failed matric or dropped out of school within the stated period. In order to ensure a basic level of comparability within the sample, all members were potentially able to matriculate by the end of 2005. This is achieved by restricting the sample to those who had completed Grade 8 - Grade 11 by the end of 2001, and were enrolled in Grade 9 - Grade 12 at the beginning of 2002.

The sample on which the Stage 2 or enrolment regressions are based is comprised of 586 YAs, all of whom passed matric between 2002 and 2005. It should be noted that Stage 2 refers only to

*enrolment* in PSE, and not completion thereof. This analysis seeks to understand the role of income and ability in getting a YA to tertiary education rather than what keeps him/her there.

The analysis focuses on Africans and Coloureds. While it would have been preferable to include Whites as well, only 103 YAs would have been eligible for the full sample, with 89 of those making it into the matric subsample<sup>7</sup>. As the focus of this study is on a racial comparison of the role of income and ability in educational attainment, the races are dealt with separately. The extremely small sample of Whites would have made their estimates unreliable. As such, Whites have been excluded entirely from the analysis.

### 3.1 Attrition

When using a longitudinal dataset it is important to account for attrition over the waves. If members of the sample drop out randomly, the effect of attrition on the estimates is unlikely to be problematic (Fitzgerald et al 1998). However, it is often the case that attrition is not entirely random, and this can introduce bias in the results. As this paper deals only with Coloureds and Africans, the attrition of Whites is irrelevant and as such will be ignored.

**Table 3.1: Attrition among Africans and Coloureds in the CAPS data, Wave 1 – 4**

	Frequency	Proportion	% Attrition
<b>Wave 1 (2002)</b>			
Africans	2148	1	0
Coloureds	2002	1	0
<b>Wave 2 (2003)</b>			
Africans	1818	0.8464	15.36
Coloureds	1690	0.8442	15.58
<b>Wave 3 (2005)</b>			
Africans	1515	0.7053	29.47
Coloureds	1676	0.8372	16.28
<b>Wave 4 (2006)</b>			
Africans	1596	0.7430	25.70
Coloureds	1591	0.7947	20.53

Table 1 shows the overall attrition rate for Africans and Coloureds between Wave 1 and Wave 4. These figures are based on the variable which captures the final result of each individual's interview. They do *not* account for the fact that some of the individuals who were not included in Wave 2, 3 or 4 were found again by Wave 5 and had their data for those years added retrospectively. In other

<sup>7</sup> The limited sample size is due to low response rates in White areas. Lam et al (2010) report that more than 80 percent of the selected Coloured and African households participated in Wave 1, as opposed to only 46 percent for White households.



words, the extent of attrition may be slightly overstated in this table due to some of these individuals being reincorporated in the survey in the 5<sup>th</sup> wave.

The table makes it clear that attrition in the sample as a whole is not racially neutral. While roughly equal numbers of Africans and Coloureds dropped out of the sample in Wave 2, the change from Wave 2 to Wave 3 is barely noticeable for Coloureds, yet much more substantial for Africans. Interestingly, by Wave 4 the total number of Africans interviewed actually increased in comparison to Wave 3, while the number of Coloureds fell by another 4 percent approximately. Nonetheless, the percentage of each race which experienced attrition is approximately 5 points higher for Africans than Coloureds by Wave 4 in 2006.

In order to assess the extent to which attrition may affect this analysis, the next table presents the average characteristics of two samples referred to as the estimation sample and the potential sample. The estimation sample includes all YAs on which this analysis is run, while the potential sample includes the estimation sample, as well as another 308 individuals who would have been

**Table 3.2: Attrition in estimation sample – average characteristics**

Variable	Estimation sample		Potential sample	
	Mean	SD	Mean	SD
<b>Observations</b>	<b>1289</b>		<b>1597</b>	
Age in 2005	20.00	1.73	20.03	1.76
Male	0.46	0.50	0.46	0.50
African	0.32	0.46	0.35	0.48
Coloured	0.68	0.46	0.65	0.48
Income	841.12	897.49	858.74	932.05
LNE score	0.16	0.98	0.14	0.99
Mother's education	8.79	2.87	8.79	2.93
Father's education	8.72	3.25	8.76	3.30
Proportion life lived with mother	0.86	0.26	0.85	0.27
Proportion life lived with father	0.59	0.44	0.58	0.44
Siblings	2.06	1.57	2.05	1.60
Grade 9 in 2002	0.28	0.45	0.28	0.45
Grade 10 in 2002	0.30	0.46	0.30	0.46
Grade 11 in 2002	0.24	0.43	0.23	0.42
Grade 12 in 2002	0.19	0.39	0.19	0.39
Pass matric 2002	0.29	0.45	0.29	0.45
Pass matric 2003	0.31	0.46	0.31	0.46
Pass matric 2004	0.22	0.42	0.22	0.42
Pass matric 2005	0.18	0.38	0.18	0.38

Notes: 1. Proportions may not sum to 1 due to rounding errors. 2. None of the means are significantly different across samples at a significance level of 5 percent.

included in the analysis had they not dropped out by Wave 4 (and failed to reappear in Wave 5). A comparison of the average characteristics across these two samples indicates the extent to which attrition may bias the estimation results.

The figures in the table reveal that the two samples are extremely similar. The proportion of Africans is slightly lower in the estimation sample and the mean LNE score is slightly higher. However, as noted below the table, none of the means are significantly different at 5 percent. Thus it appears that for the purpose of this analysis, the attrition which occurred among the individuals of interest is unlikely to have any effect on the estimation results.

## 3.2 Descriptive statistics

Table 3.3 provides an overview of the average characteristics of the full sample, the subsample with matric and the subset thereof that enrolled. The first point to note is that the racial composition changes slightly as the sample under consideration becomes more educated. Africans make up 32 percent of the full sample, but only 28 percent and 27 percent of the matriculated and enrolled subsamples respectively. The change in racial composition is not dramatic but it does indicate a slight advantage held by Coloureds over Africans in terms of educational attainment.

Of most relevance to this study is the pattern in manifest ability (LNE score) and income across samples. As the LNE score has been normalised by age category within the full sample, the numbers themselves do not have much meaning; it is the relative scores which are important. It is clear that the LNE score, or ability, rises as the sample becomes more educated. The standard deviation of the LNE score for the full sample is just less than one (when age is not taken into consideration and weights are applied). This means that the mean LNE score rises by slightly more than 0.40 standard deviations when you move from the full sample to the subsample of matriculants. The jump up is smaller, but still notable, when moving from all matriculants to only those who enrol – just over 0.21 standard deviations.

A relatively large increase in per capita household income across samples is also evident. While the average for the full sample stands at R841.12 in 2002, it is approximately 29 percent greater for those who matriculate, at R1087.69, and a further 28 percent greater for those who enrol, at R1388.90. In contrast to the pattern with ability, the jump up in mean income between the full sample and the matric subsample is more or less equal to that between the matric and enrolled

subsample. This could be the first indication that while both ability and income are positively correlated with educational attainment, their bearing on this process differs across the stages.

**Table 3.3: Means and standard deviations by level of education**

Variable	Full sample		Matriculated sample		Enrolled sample	
	Mean	SD	Mean	SD	Mean	SD
<b>Observations</b>	<b>1289</b>		<b>586</b>		<b>259</b>	
Age in 2005	20.00	1.73	19.88	1.58	19.87	1.45
Male	0.46	0.50	0.45	0.50	0.48	0.50
African	0.32	0.46	0.28	0.45	0.27	0.45
Coloured	0.68	0.46	0.72	0.45	0.73	0.45
Income in 2002	841.12	897.49	1087.69	1067.65	1388.90	1258.32
LNE score	0.14	0.99	0.53	0.89	0.73	0.89
Mother's education	8.79	2.87	9.44	2.86	10.38	2.80
Father's education	8.72	3.25	9.56	3.17	10.45	3.09
Proportion life lived with mother	0.86	0.26	0.89	0.24	0.91	0.22
Proportion life lived with father	0.59	0.44	0.64	0.43	0.67	0.43
Number of siblings	2.06	1.57	1.95	1.45	1.73	1.32
Matriculated	0.48	0.50	1	0	1	0
Enrolled   Matric	0.23	0.42	0.47	0.50	1	0
Grade 9 in 2002	0.27	0.45	0.14	0.34	0.13	0.34
Grade 10 in 2002	0.30	0.46	0.24	0.42	0.20	0.40
Grade 11 in 2002	0.24	0.43	0.32	0.47	0.32	0.47
Grade 12 in 2002	0.19	0.39	0.30	0.46	0.35	0.48
Matriculated in 2002	0.29	0.45	0.29	0.45	0.34	0.47
Matriculated in 2003	0.30	0.46	0.30	0.46	0.30	0.46
Matriculated in 2004	0.23	0.42	0.23	0.42	0.19	0.40
Matriculated in 2005	0.18	0.39	0.18	0.39	0.17	0.37

Notes: 1. The data are weighted to account for sample design and individual non-response. 2. Due to the exclusion of observations with missing information on mother's and father's education, the samples on which these figures are calculated are 1170 and 824 respectively.

An examination of the other variables indicates the extent to which educational attainment is related to family characteristics. The stronger the relationship, the more important it is to account for these factors when examining the role that income and ability play in matriculation and enrolment.

The variables capturing family characteristics include parental education, proportion of life lived with each parent until Wave 1 in 2002 and the number of siblings. In terms of parental education, it is clear that the average number of years of schooling increases for both parents as the sample of YAs becomes more educated. For example, both mothers and fathers of the enrolled subsample hold almost two more years' schooling on average than parents of the full sample. This is in line with much of the literature which finds that parental education is positively correlated with children's

education. The proportion of life lived with each parent also increases as the sample becomes more educated, although the increase does not exceed 8 percentage points for either parent. It should be noted that for all three samples, the proportion of a YA's life lived with their father is approximately 25% lower than with their mother. Calculations across the three samples (not shown here) reveal that between 25-29 percent of YAs have *never* lived with their father while the equivalent figure for mothers is only 1-2%. This reflects the high prevalence of single parent families among non-Whites. As the literature indicates that this has a negative impact on educational outcomes, it should be borne in mind when looking at matriculation and enrolment rates. The final family variable is the number of siblings, which decreases on average as the sample becomes more educated.

The last eight variables report the proportion of the sample enrolled in Grades 9 - 12 in 2002, as well as the proportion of matriculants who passed Grade 12 in the years 2002 – 2005. These variables are included in the econometric analysis in order to control for the way in which the sample has been constructed. For those enrolled in Grade 12 in 2002, there are four opportunities to matriculate, while those in Grade 9 only have one opportunity. An individual is included in the tertiary sample if they enrol between 2003 and 2006 (inclusive), indicating that those who matriculate in an earlier year have more chances to enrol than those who matriculate in a later year. Dummy variables which capture grade of enrolment and year of matriculation are therefore included to control for the fact that YAs who are more advanced in the schooling system will be more likely to matriculate and enrol regardless of their other characteristics.

The change in the proportions across samples makes it clear that having more years in which to matriculate and then enrol does affect the probability of these events happening. The proportion of the sample made up by those enrolled in Grade 9 and 10 in 2002 decreases as the sample becomes more educated, while the proportion of the sample comprised of those enrolled in Grade 11 and 12 rises. A similar but more subtle pattern is evident for the year of matriculation variables: those who matriculated in 2002 make up a greater proportion of the enrolled sample than the matriculated sample, while the opposite is true for those who matriculated in 2004 or 2005.

It is clear from this overview that most of the included characteristics are related and move together with income and ability in explaining educational attainment. Those who manage to matriculate have both higher ability and higher income, they have better educated parents and have spent a greater proportion of their life living with their parents. Out of the matric subsample, those who enrol also fare better on average for all of these factors. This pattern points to the necessity of controlling for these factors when attempting to determine the role that income and ability play in predicting educational attainment.

Before turning to kernel densities and lowess estimates which focus specifically on income and ability, it is useful to have a brief look at the racial means in the full sample, which are shown in Table 3.4. Understanding how Africans and Coloureds vary in terms of all characteristics places each race's relationship between income, ability and educational attainment in context.

**Table 3.4: Means and standard deviations by race for the full sample**

Variable	Africans		Coloureds		Sig. at 5%
	Mean	SD	Mean	SD	
<b>Observations</b>	<b>616</b>		<b>673</b>		
Age in 2005	20.98	1.91	19.56	1.43	*
Male	0.44	0.50	0.46	0.50	
Income in 2002	372.96	366.24	1056.91	982.86	*
LNE score	-0.32	1.00	0.35	0.91	*
Mother's education	8.21	3.10	9.04	2.72	*
Father's education	7.56	3.47	9.16	3.05	*
Proportion life lived with mother	0.78	0.29	0.90	0.23	*
Proportion life lived with father	0.44	0.41	0.66	0.43	*
Siblings	2.63	1.85	1.80	1.34	*
Matriculated	0.42	0.49	0.50	0.50	*
Enrolled   Matric	0.46	0.50	0.48	0.50	
Grade 9 in 2002	0.28	0.45	0.27	0.45	
Grade 10 in 2002	0.28	0.45	0.30	0.46	
Grade 11 in 2002	0.25	0.43	0.24	0.42	
Grade 12 in 2002	0.19	0.39	0.19	0.39	
Matriculated in 2002	0.24	0.43	0.31	0.46	
Matriculated in 2003	0.29	0.46	0.32	0.47	
Matriculated in 2004	0.22	0.42	0.22	0.42	
Matriculated in 2005	0.25	0.43	0.15	0.35	*

Note: 1. The data are weighted to account for sample design and individual non-response. 2. Due to the exclusion of observations with missing information on mother's and father's education, the sample on which these figures are calculated is 1170 and 824 respectively.

The figures make it clear that the races differ on most bases assumed to play a role in educational attainment. The mean LNE score for Africans is almost 0.7 standard deviations below that of Coloureds. Africans live on approximately a third of the per capita monthly income of Coloureds, and both parents are less educated on average than Coloured parents. Furthermore, African YAs have lived with both parents for less of their lives (until the age of 14), and they have more siblings. These differences are all significant at 5 percent. Given that these findings reveal that Africans come from relatively disadvantaged backgrounds, it is interesting that the proportion of Coloureds with matric exceeds that of Africans by only 8 percent, and that the enrolment rate (conditional on matriculation) is not significantly different between the races.

Overall these basic statistics indicate that what is known to be true for South Africa as a whole is true within this sample: Africans fare worse than Coloureds on many of the factors which comprise socio-economic wellbeing. The fact that this does not translate into substantially different educational outcomes hints at the possibility of a different process of educational attainment occurring within each group.

Figures 1 - 8 display the lowess estimates for matriculation and enrolment by income and LNE score for Africans and Coloureds, along with the kernel densities of income and LNE score for the relevant sample in the lower panel. Note that the densities exclude the bottom and top one percentile of observations in order to ignore outliers. Dashed lines have been added to each graph to highlight the area of common support, or overlap, between the two races.

The idea of common support is important and deserves to be expanded on slightly. While the focus of research and political debate in South Africa is often on the persistence of inequality between Whites and non-Whites, it is also true that substantial differentiation exists *within* non-Whites. The kernel densities in this chapter reveal this differentiation by showing the extent to which Coloureds hold the upper hand in terms of income and manifest ability. In each of the density graphs plotted in Figures 1-8, the Coloured distribution lies distinctly to the right of the African distribution. At the same time, it is obvious that the races do overlap to a large degree. If one is to compare the races in a meaningful way, as is the attempt in this paper, there must be some area of common support. In other words, there should be sufficient values of income and LNE score at which the outcome of interest is observed for both races. While the comparison of Africans and Whites can be problematic because this area of common support is small (see Figures 1 and 2 in Lam et al 2010: 27-28), it appears that there is sufficient overlap between Africans and Coloureds to make the comparison valid.

In terms of the calculation, the common support numbers reported at the bottom of each kernel density include YAs who have a value for income or LNE score which lies between the lower bound of the Coloured distribution and the upper bound of the African distribution. This means that Africans who report a value for income or LNE score which lies below the Coloured minimum are excluded, as are Coloureds who report values above the African maximum .

The kernel densities in Figures 2 and 4 indicate the disparity between Africans and Coloureds in terms of per capita household income. While the densities do overlap, it is nonetheless true that Coloureds hold the upper hand in terms of financial well-being.

Figure 1  
Proportion matriculated 2002-2005, by income

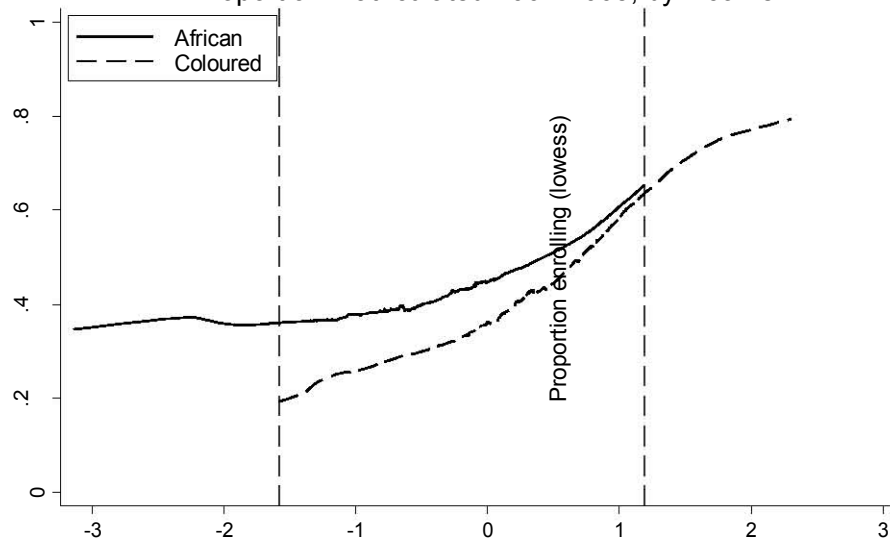


Figure 3  
Proportion enrolled 2003-2006, by income

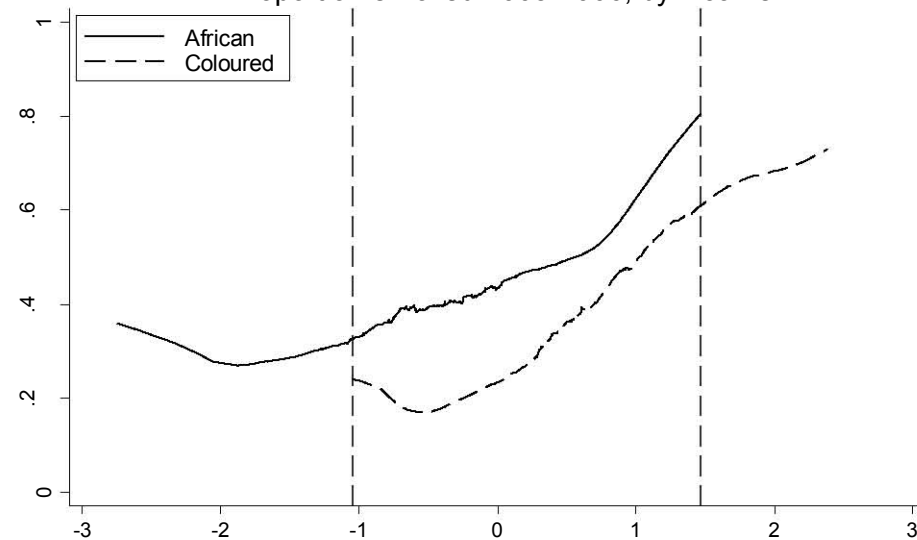
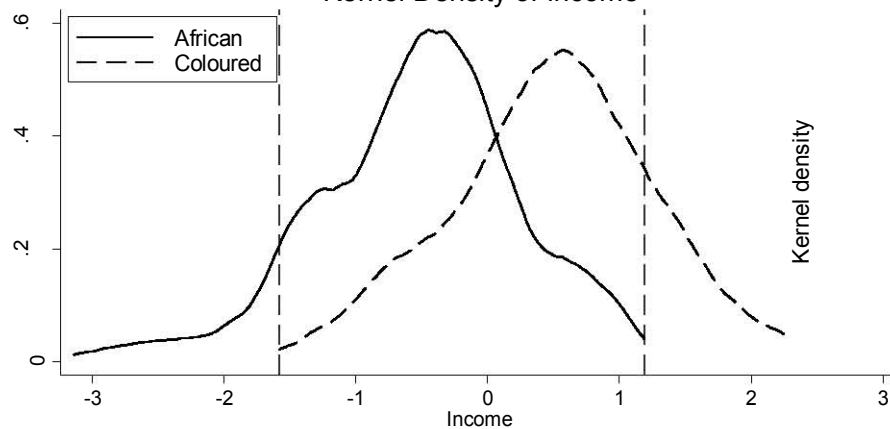


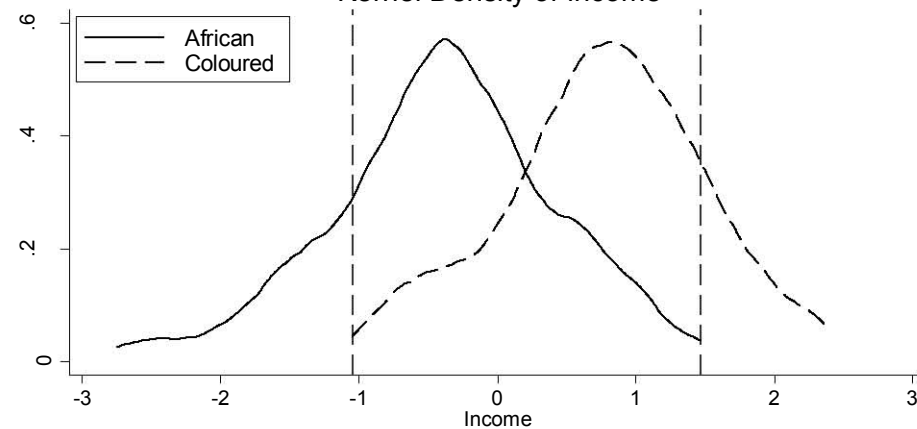
Figure 2  
Kernel Density of income



Common support  
A:547  
C:533

Epanechnikov kernel with plug-in bandwidth

Figure 4  
Kernel Density of income



Common support  
A:200  
C:254

Epanechnikov kernel with plug-in bandwidth

Turning to the lowess curves which estimate the probability of matriculation or enrolment over the income distribution, it is clear that Africans are more likely to both matriculate and enrol within the area of common support (bearing in mind that other characteristics are *not* held constant here). Interestingly, the gap between the races' lowess curves is greater for enrolment than matriculation, implying that the African advantage over Coloureds would be greater at the post-secondary level if income is held constant.

Examining the slopes of the lowess estimates is one way to determine the sensitivity of matriculation and enrolment to income. For both stages of educational attainment, the lowess curve is steeper on average for Coloureds than Africans. This indicates that Coloureds are more sensitive to a change in income than Africans. Taken at face value, if a monthly transfer of R1000 were given to a group of each race, we might expect to see the proportion matriculating and enrolling to rise by more within the Coloured group than the African group. In addition, it appears that the slopes for each race increase from Figure 1 to Figure 3, pointing to a greater sensitivity to income at the enrolment stage for both races.

Figures 5 - 8 repeat the previous four figures with regards to the LNE score. The kernel densities once again reveal that Coloureds have an advantage over Africans in terms of their distribution of manifest ability.

The lowess estimates present some interesting findings. In Figures 5 and 7, it appears the races have similar slopes at lower levels of ability. However, past a threshold (just above zero in Figure 5 and one in Figure 7), the slope increases for the Coloured lowess curve while remaining constant (Figure 5) or decreasing (Figure 7) for the African curve. Thus it could be that the proportion of Coloureds matriculating or enrolling responds more to a change in ability when ability is initially higher, but the same is not true for African YAs.



Figure 5  
Proportion matriculated 2003-2006, by LNE

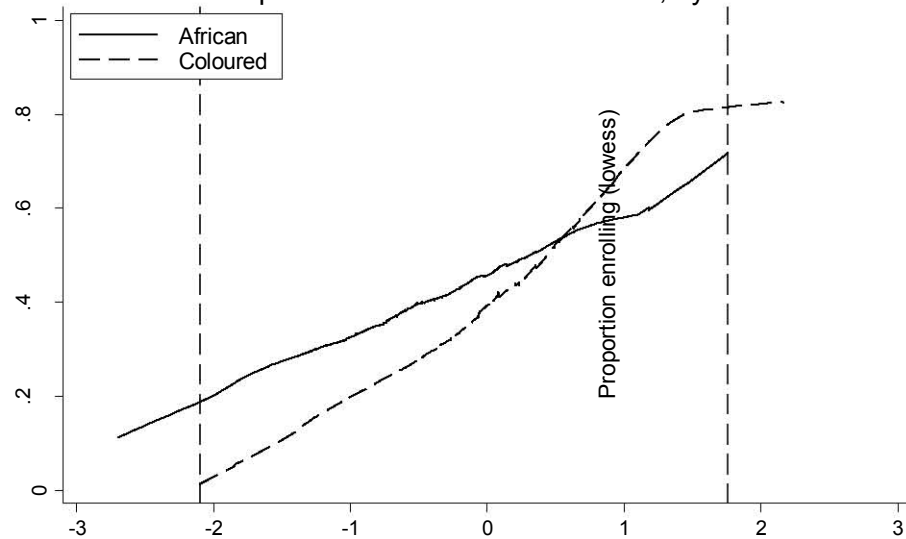


Figure 7  
Proportion enrolled 2003-2006, by LNE

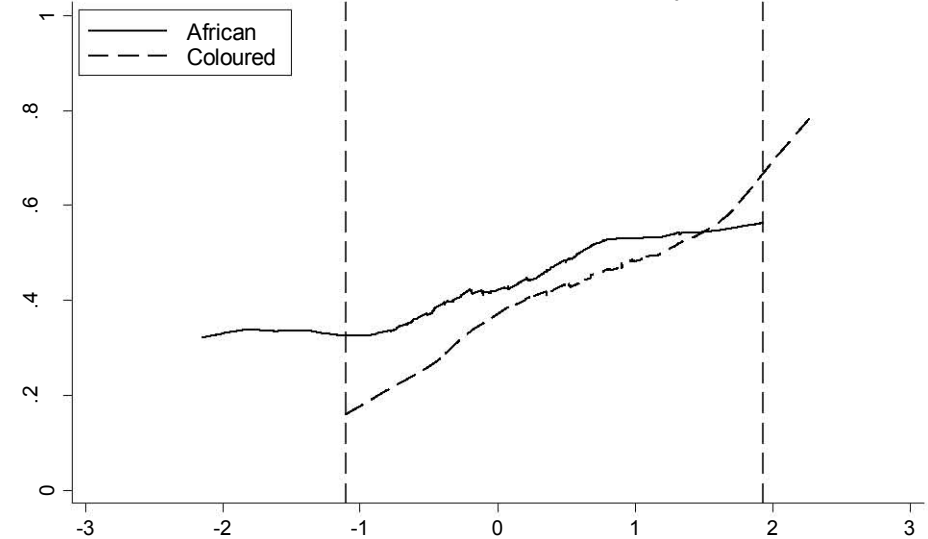
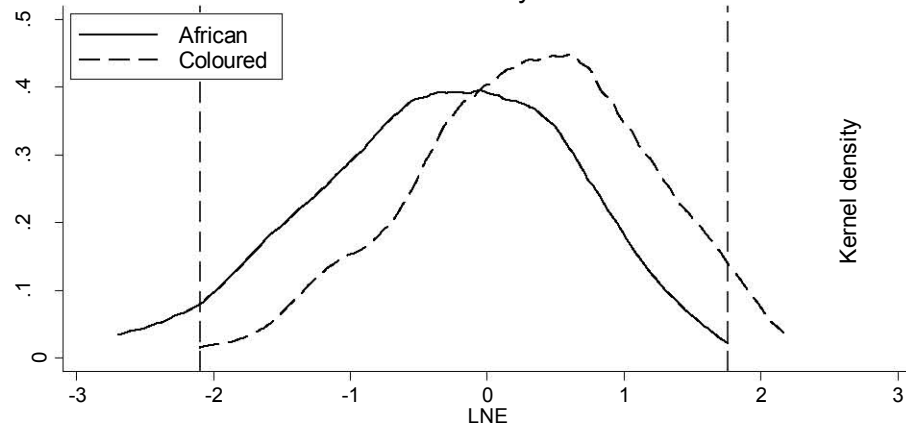


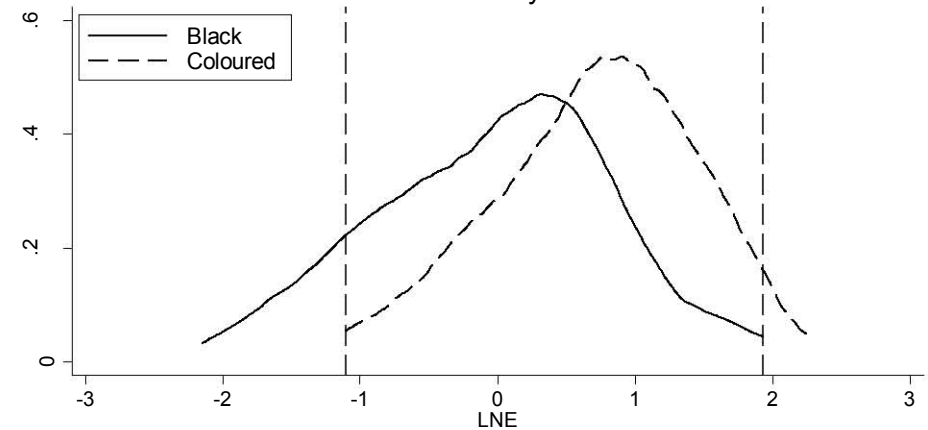
Figure 6  
Kernel density of LNE



Common support  
A:583  
C:641

Epanechnikov kernel with plug-in bandwidth

Figure 8  
Kernel density of LNE



Common support  
A:223  
C:312

Epanechnikov kernel with plug-in bandwidth

A difference between this second set of lowess estimates and those in Figures 1 and 3 is that the proportion of Coloureds matriculating or enrolling surpasses that of Africans at higher LNE scores, whereas Africans hold the advantage when the races were compared over shared income values. Equalising income makes Africans more likely to attain education at both stages than Coloureds, while equalising ability only gives Africans the advantage at the lower end of the distribution.

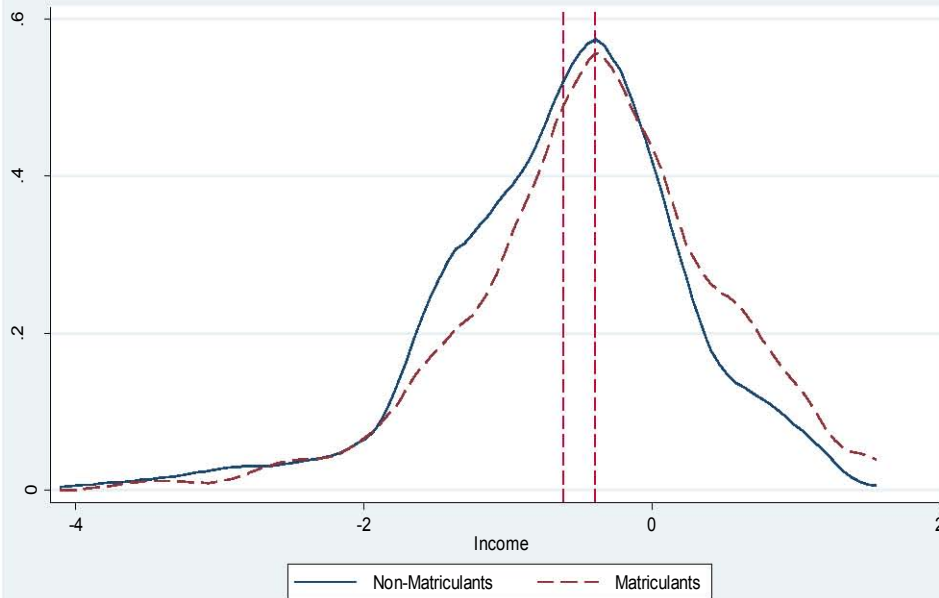
Figures 9 - 12 plot the kernel densities of income and LNE score for matriculants and non-matriculants, by race. While the previous kernel densities served to highlight the disparity between races, these kernel densities illustrate the extent to which members within a race group differ according to their educational status. Note that the mean income and LNE score for both densities are reported below the graph, as well as whether or not the difference in means is significant at 5%.

Figures 9 and 10 are kernel densities of income for Africans and Coloureds respectively. For both races, mean income is significantly higher for matriculants than non-matriculants. With this being said, the gap in mean income is larger for Coloureds, and the distribution of matriculants' income is shifted further to the right than for the equivalent African density.

The kernel densities for LNE score are displayed in Figures 11 and 12. The findings are similar to those for Figures 9 and 10. Matriculants within both races have a significantly higher average LNE score than non-matriculants. Furthermore, the difference in mean LNE score is greater for Coloureds than Africans, and the Coloured matriculant distribution is shifted further to the right. These findings all point to a greater sensitivity to both income and ability among Coloureds than Africans, as was indicated by the slopes of the lowess curves in Figures 1 and 3.

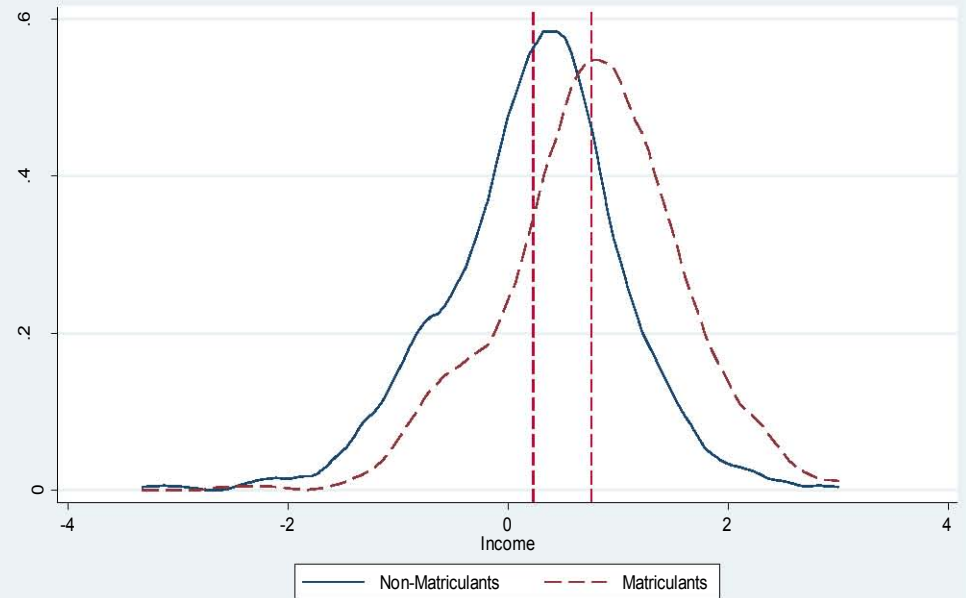
## Kernel density of income Matric versus Non-Matric

Figure 9  
Africans



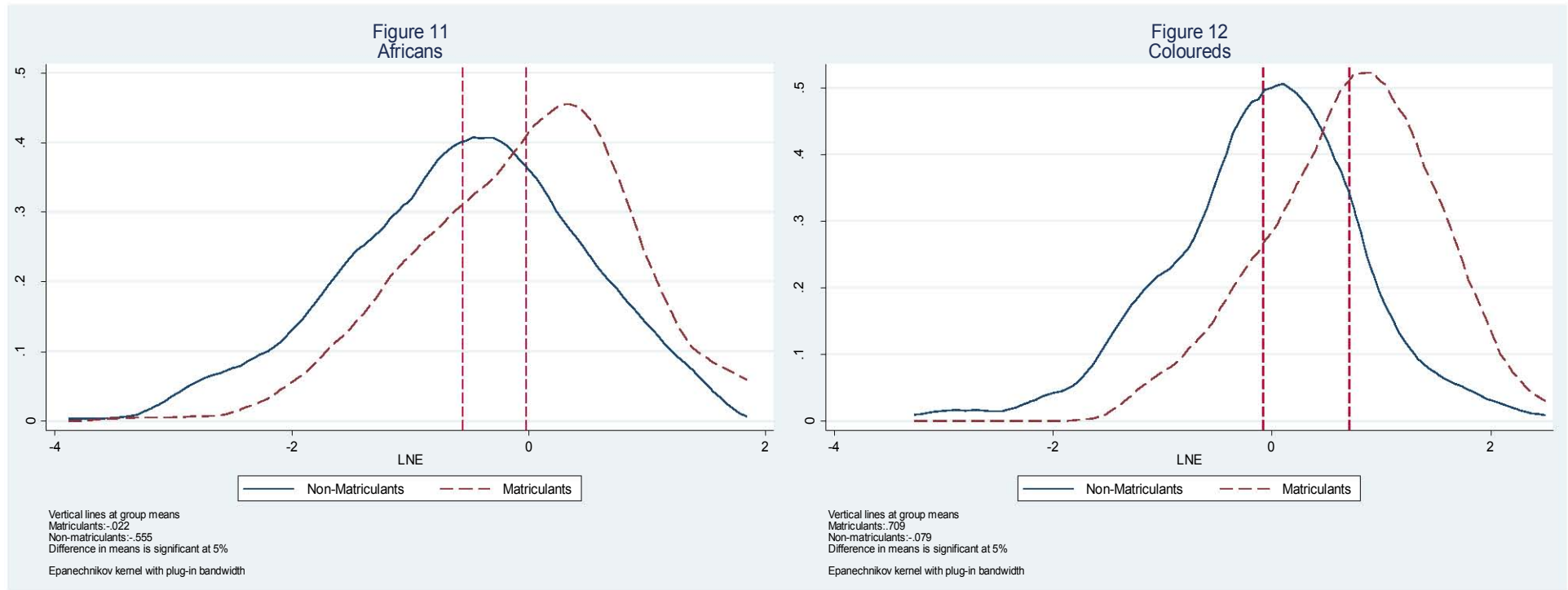
Vertical lines at group means  
 Matriculants: -.392  
 Non-Matriculants: -.614  
 Difference in means is significant at 5%  
 Epanechnikov kernel with plug-in bandwidth

Figure 10  
Coloureds



Vertical lines at group means  
 Matriculants: .761  
 Non-Matriculants: .236  
 Difference in means is significant at 5%  
 Epanechnikov kernel with plug-in bandwidth

## Kernel density of LNE Matric versus Non-Matric

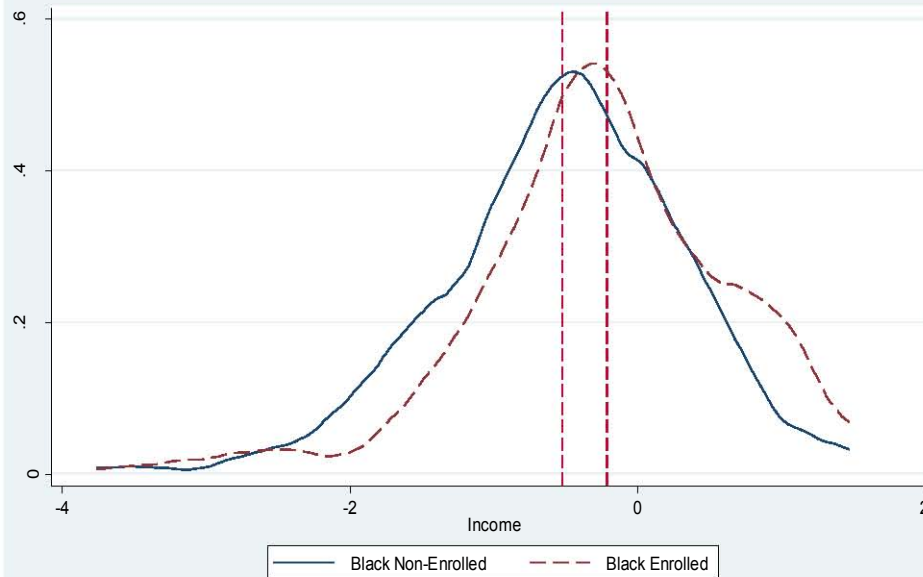


Figures 13 – 16 repeat the previous four graphs for the enrolment stage. Figures 13 and 14 reveal that average income is higher for those who enrol as compared to those who do not (given that all YAs included in these densities have matriculated). However, the difference is not statistically significant for Africans, while it is for Coloureds. This is not surprising, as visual inspection makes it clear that the gap between the average incomes of the two groups is relatively small for Africans, while being much greater for Coloureds. Furthermore, the African distributions are largely overlapping while the Coloured distribution of enrolled YAs lies noticeably to the right of the non-enrolled YA distribution.

The final two densities, Figures 15 and 16, follow the pattern of the previous two figures. The mean LNE score is higher for enrolled YAs than non-enrolled YAs, but the difference is significant only for Coloureds. The African distribution for enrolled YAs does appear to be shifted further to the right than the equivalent density for income, suggesting that perhaps ability is more differentiated between enrolled and non-enrolled Africans than income. Again, Coloureds display a substantial and significant difference in mean LNE score between the two groups, as well as a noticeable shift to the right of the enrolled distribution.

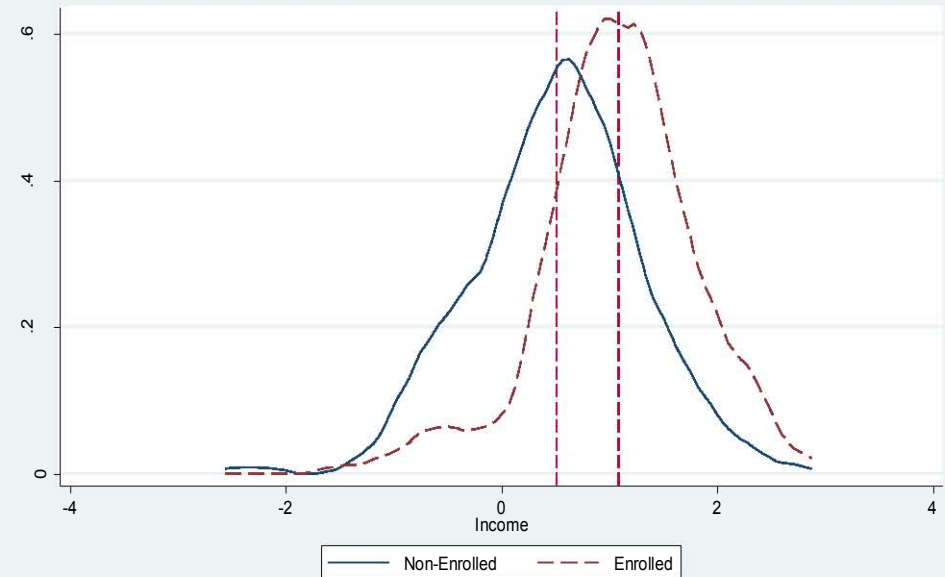
## Kernel density of income Enrolled versus Non-Enrolled

Figure 13  
Africans



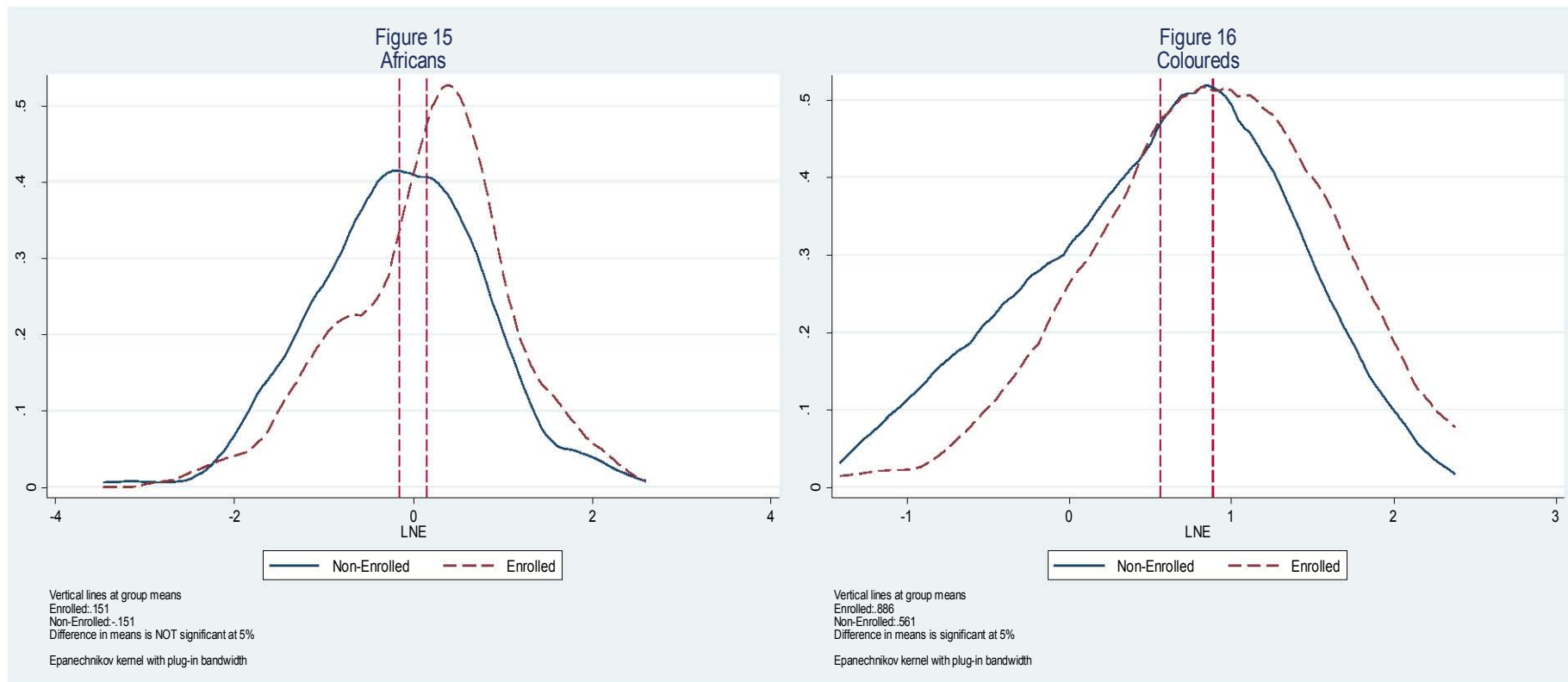
Vertical lines at group means  
Enrolled:-.208  
Non-Enrolled:-.519  
Difference in means is significant at 5%  
Epanechnikov kernel with plug-in bandwidth

Figure 14  
Coloureds



Vertical lines at group means  
Enrolled:1.087  
Non-Enrolled:.516  
Difference in means is significant at 5%  
Epanechnikov kernel with plug-in bandwidth

## Kernel density of LNE Enrolled versus Non-Enrolled



As well as highlighting the differences within groups, the previous eight kernel densities provide further evidence of the gap between races. For example, the average income for African matriculants is significantly and substantially lower than Coloured *non*-matriculants. The same is true of enrolled Africans as compared to non-enrolled Coloureds. With respect to ability, African matriculants and Coloured non-matriculants have roughly the same mean LNE score, while Africans who enrol fare significantly worse than Coloureds who do not enrol. Thus it seems that while the probability of matriculation and enrolment rises as income and ability rise for both Africans and Coloureds, this process is happening at almost entirely different levels for the two races.

The comparison of the races over the area of common support in Figures 1 – 8 provided evidence of this phenomenon. While the statistics in Table 3.4 indicated that Coloureds have a significantly higher rate of matriculation than Africans in reality, the lowess curves revealed that Africans are more likely to both matriculate and go on to enrol when income is held constant. This is also true when ability is held constant (over the majority of shared LNE scores). Thus it seems possible that the educational advantage held by Coloureds is due largely to their superior income and ability distributions. If, on average, Africans are managing to matriculate and enrol at levels of income and ability where Coloureds are failing to do so, then it is not surprising that Africans would be estimated to do better when these variables are held constant. The analysis in Chapter 5 will build on this finding.

Regarding the nature of the relationship between the two variables of interest, income and ability, and the two stages of educational attainment, matriculation and enrolment, several ideas can be drawn from the preceding descriptive examination. For both races, both income and ability are significantly higher for those who matriculate as opposed to those who do not. Furthermore, the difference in means and distributions between matriculants and non-matriculants is greater for Coloureds than for Africans with regards to both income and ability. Thus it appears that both factors are important in predicting matriculation, and may be of greater importance for Coloureds than Africans. Looking at the second stage, neither mean income nor mean ability is significantly higher for Africans who enrol as opposed to Africans who do not enrol. In contrast, both are significantly higher for enrolled as opposed to non-enrolled Coloureds. In terms of the distributions, it appears that ability may play some role in predicting enrolment for Africans despite the insignificance in means. The analysis will seek to quantify these relationships and provide an idea of the difference in the relationship across stages and between races.



## 4. METHODOLOGY

While much has been learnt from the lowess estimates and kernel densities, it is necessary to use multivariate modeling in order to disentangle the complicated social processes which are driving these outcomes. These will be tailored to address the research questions at hand; namely, what role do income and manifest ability play at two important stages: obtaining matric, and going on to enrol in PSE. In order to provide as clear a picture of these processes as possible, income and ability will be analysed separately, with a third section dealing with the interaction of the two variables.

Several different techniques will be used to assess the role of income and ability. Marginal effects and discrete changes in probability are presented by race. These estimates are reflections of the sensitivity of each race's educational attainment to a change in income or ability. Graphs of the predicted probability of matriculation and enrolment are also presented, which allow a racial comparison of educational attainment over the entire income and ability range. How these methods will be implemented is explained in this chapter.

### 4.1 Marginal effects

The first step in the analysis is to run a set of three logit regressions<sup>8</sup> for matriculation, and enrolment given matriculation (also referred to as conditional enrolment), progressively adding sets of variables from the first to the third equation. As this study is specifically interested in a comparison of the processes of educational attainment between Africans and Coloureds, the races are analysed separately from the outset. This is in line with Cameron and Heckman (2002), as well as Lam et al (2008).

In all regressions, age, age squared and gender are included. In addition, dummy variables which account for the grade of enrolment in 2002 are included in the matriculation regression, while the enrolment equation incorporates dummy variables for the year of matriculation. In the first regression, only the variable of interest is included (i.e. income or ability in their corresponding sections) in addition to the aforementioned variables. Regression 2 adds the other variable of interest (LNE score in the income section and vice versa). The third and final regression incorporates seven variables relating to family characteristics: mother's education, father's education (both captured as continuous variables), number of siblings, proportion of life lived with mother/father until the age of 14 and two dummy variables capturing missing information on parental education.

---

<sup>8</sup> A brief derivation of the logit model is provided on page 38, and as such is not dealt with here.

Adding variables progressively in this way is done for two reasons. Firstly, in line with Carneiro and Heckman (2002), it reveals the extent to which income remains significant as a predictor of educational attainment when individual and family characteristics are accounted for. This analysis is interested in the importance of *current* income (or credit constraints) in predicting educational attainment. Without accounting for individual ability and family characteristics, the income variable is likely to pick up the effects of both current and long-term income. The latter determines factors such as household resources and quality of schooling, among others. It is necessary to account for those factors separately, in order to isolate the effect of current income.

Secondly, the addition of background variables plays a similar role for the LNE score. As discussed in the literature review, this study is interested in manifest and not innate ability. The inclusion of income and family background is not an attempt to purge the LNE score of their effects. Such a process would only be appropriate if a measure of innate ability is desired. Moreover, this would require a separate regression of the LNE score on these and other variables (see Branson et al 2009). Rather, their inclusion is to ensure that the coefficient on the LNE score does not pick up the *direct* effects of household resources and parental quality on matriculation and enrolment rates (which do not operate via the youth's ability).

It should be noted that the LNE score used in these regressions is standardised by the age at which the evaluation was taken. In practical terms, this involves subtracting the mean LNE score for a respondent's age group from the respondent's raw score, and dividing it by the standard deviation of scores within that age category. This results in a score which has a mean of 0 and a standard deviation of 1 within each age category in the full sample. This transformation accounts for the tendency of the LNE score to increase with age, given that it is determined by acquired skills and knowledge in addition to innate ability<sup>9</sup>. As grade failure means that progression through school is not always perfectly correlated with age, the grade of enrolment in 2002 is also included in all matriculation regressions as an additional adjustment to account for prior learning.

The coefficients produced by a logit regression are problematic to interpret. They represent the change in the log odds of an event occurring, which is not intuitively easy to understand. By taking the log of the coefficients we obtain the change in the odds for a given change in the corresponding variable, which is simpler. However, the easiest way to understand logit coefficients is to transform them into marginal effects, which are the partial derivative of the outcome with respect to a given independent variable.

---

<sup>9</sup> Belley and Lochner (2007) implement a similar adjustment in order to account for the observed positive correlation between age and test scores used to measure cognitive ability.

Long (1997) explains that marginal effects are generally calculated in two slightly different ways. The first method involves taking the partial derivative of the predicted probability with respect to a given independent variable<sup>10</sup>, while holding all other variables at their mean. This is called the marginal effect at the point of means. In contrast, the average marginal effect is calculated by averaging the marginal effect of a given variable across all individuals. The second method potentially represents a more useful measure because the point of means may not characterise a feasible person in reality (Long 1997). However, it is common practice in much of the literature to present the marginal effects at the point of means and it is this method which will be used in the analysis section.

## 4.2 Predicted probabilities

While marginal effects are a useful way to determine the relationship between a given variable and the outcome, they potentially have a weakness. As Long (1997) explains, they may under- or overestimate the true change in probability due to the method of calculation. Taking a partial derivative involves estimating the slope of a tangent to the curve at a particular set of values. If the relevant section of the curve is non-linear, a tangent is at best an approximation of the true slope at that point (1997: 75).

A more serious constraint imposed by marginal effects relates to the comparison of coefficients across groups. Long (2009) argues that “traditional tests of the equality of regression coefficients across groups confound the magnitude of the regression coefficients with residual variation” (2009: 11). Instead, Long suggests using predicted probabilities when studying group differences. A full derivation and explanation of Long’s viewpoint can be found in his article. Here I will attempt to relay his argument and apply it to this example.

A good point at which to start is with a brief derivation of the logit model. This serves to highlight the difference between linear and non-linear models which makes direct comparison of coefficients appropriate in the former but inappropriate in the latter. Underlying the logit is the idea of a latent continuous variable, usually referred to as  $\eta$ . In this case,  $\eta$  would be the propensity to matriculate and following that, to enrol. However, this variable is termed latent because it is not observed. Instead, we only observe whether or not an individual does actually matriculate or enrol. If we assume the model for the latent variable is

---

<sup>10</sup> If the independent variable is continuous then the change in  $x$  is infinitesimal, while if the variable is discrete the change is from zero to one.

where  $\mathbf{x}$  represents a vector of independent variables, then the relationship between  $y$ , the observed outcome, and  $\eta$ , the latent variable, can be represented as

Thus there is a threshold, usually assumed to be zero for simplicity's sake, which links  $y$  to  $\eta$ . Applied to this paper, the relationship indicates that once the propensity to matriculate/enrol has exceeded the threshold, a YA is observed to matriculate/enrol. For all points where the propensity is equal to or below the threshold, a failure to matriculate/enrol is observed.

The point of divergence between linear and non-linear models which is important in this analysis is the way in which the variance of the error term,  $\sigma^2$ , is treated. In linear models, the variance of the error term can be estimated because  $y$  is observed. However, in non-linear probability models,  $\eta$  is not observed and hence the variance cannot be estimated and instead must be assumed. This is an identification assumption, because the model cannot be solved without it. In the logit model, the variance is arbitrarily assumed to be  $\frac{\pi^2}{3}$ .<sup>11</sup>

This identification assumption does not present problems until a direct comparison of group coefficients is attempted. In this analysis, the groups are races. When Africans and Coloureds are dealt with in separate regressions, two models are implicitly proposed:

Africans:

Coloureds:

In order to force the variance of the error term to equal  $\frac{\pi^2}{3}$  in both equations, they must be transformed in the following way:

$$\eta = \frac{y - \beta_0 - \beta_1 x}{\sigma} \quad \eta = \frac{y - \beta_0 - \beta_1 x}{\sigma}$$

<sup>11</sup> Long's example refers to the probit, but he states that his findings can be applied equally to either model. I refer only to the logit as that is the model used in the analysis.

$$\beta_1 = \beta_2 = \beta_3 = \beta_4$$

If one uses standard statistical software, such as Stata, to test the equality of regression coefficients, the hypothesis which will be tested is

$$\beta_1 = \beta_2$$

or, put more simply,

$$\beta_1 - \beta_2 = 0$$

However, Stata cannot test the hypothesis,

As  $\beta_1$  includes only the coefficients of interest, leaving out the residual variation in each group, it is the correct hypothesis for a group comparison.  $\beta_1 = \beta_2$  and  $\beta_1 - \beta_2 = 0$  are only equivalent if it is assumed that  $\sigma^2 = 0$ . However, there is no a priori reason for this assumption to be correct. Thus a direct test of the equality of two regression coefficients confuses the actual coefficients with the residual variation present in each group, and can lead to incorrect conclusions.

Long's solution to this conundrum is to compare groups using predicted probabilities as they are unaffected by assumptions with regard to the variance of the error term. While the transformation depicted above only allows a scaled version of the coefficients to be compared across groups, the predicted probability for a given vector of  $X$  is unaffected by such scaling. In order to see this, consider the following equation

This equation phrases the previously stated relationship between  $\beta_1$  and  $\beta_2$  in terms of probability. The right hand side can then be rewritten in terms of the error,

Applying this to the African model once the identification assumption has been imposed, the probability of an African YA matriculating or enrolling can be written as

$$\beta_1 = \beta_2 = \beta_3 = \beta_4$$

As the scaling factor of  $\frac{1}{\sigma}$  appears in every term on the right hand side, the expression can be simplified to

In other words, while the transformation causes the coefficients to be identified only to the point of their scaled value, it does not affect the predicted probabilities for a given  $x$ . This means that it is valid to compare predicted probabilities across groups.

### 4.2.1 Discrete changes in predicted probability

Predicted probabilities are used in two ways in this study. Firstly, the discrete change in the predicted probability of educational attainment due to a standard deviation change in income and ability around their respective means is presented for each of the three regressions. Each race is analysed separately, allowing a comparison of the discrete change to be compared across population groups. This takes the place of directly testing the equality of marginal effects across races, and thus avoids the problem of residual variation.

The calculation proceeds as follows

$$\frac{\partial \hat{P}_i}{\partial x} \sigma_x$$

where  $\sigma_x$  represents a standard deviation change in income or ability (Long 1997: 77). A standard deviation is used in order to address the fact that income and ability are measured on different scales. A comparison is only meaningful if the increase in each variable is proportional and using a standard deviation ensures that this is the case.

This discrete change is calculated for matriculation and enrolment conditional on matric. The reason for analysing *conditional* enrolment relates to the nature of educational attainment in South Africa. A number of papers which deal with enrolment into tertiary level education in South Africa restrict the sample to those with matric (e.g. Branson et al 2009, Lam et al 2010). Framing the process of educational attainment in this way suits the South African context to a large degree, as most (but not all) tertiary institutions require a matric certificate in order to enrol. This is reflected in the CAPS data, which includes a question asking if matric is a pre-requisite for enrolment in the respondent's current PSE institution. In the years 2004 – 2006, between 82 and 88 percent of the respondents within the full dataset indicate that their current institution requires matric. Note that this

calculation is *not* restricted to those with matric. Moreover, only 20 people within the estimation sample report being enrolled in some form of PSE without holding matric, as opposed to 259 individuals who do have matric. It therefore makes sense to analyse enrolment conditional on matric as it ensures that the members of the included sample are on as equal a footing as possible with regards to the pursuit of further education.

However, it should be borne in mind that predicted probability changes derived from a conditional enrolment regression only estimate the impact that income and ability have on tertiary education over and above the impact already felt at the matriculation stage. It is likely that YAs who are eligible to enrol in PSE are a select sample of all YAs who attend secondary school. They have either been selected on the basis of observables, such as income, ability and family background, on the basis of unobservables such as determination and discipline, or on some combination of the two. Whatever the process of selection, it is highly unlikely that those who matriculate are statistically the same on all bases as those who fail to matriculate. This is supported by Table 3.3 in Chapter 3, which demonstrated that the full sample, the matric subsample and the enrolled subsample do indeed differ on observables, such as income, LNE score, parental education, parental co-residence and number of siblings.

Given this process of selection, the conditional estimates are likely to hold only for matriculants, and cannot be generalised to all YAs. The unconditional probability of enrolment provides a way to calculate an estimate that can be generalised beyond this subset of YAs. Two methods are used to estimate the effect each of these variables has on the unconditional probability of enrolment.

The first method is the simpler of the two, and involves estimating the enrolment regression on the full sample of YAs rather than the subsample with matric. This regression avoids the problem of selection bias by including the entire sample in the regression. However, it implicitly assumes that any youth who reaches high school may enrol in tertiary education, and therefore ignores the sequential nature of educational attainment in South Africa. As such, it could provide biased estimates. It is included as a counterpoint against which to compare the conditional estimates, as well as the estimates generated using the second method.

The second method uses Bayes' Law, and is taken from Maddala (1983). The author considers sequential response models, using the same case of educational attainment as his example. His outcomes are phrased in the negative rather than the positive (an individual has *not* finished high school as opposed to he has finished high school), and he considers four stages rather than just two. Aside from these differences, his methodology can be applied directly to this analysis. Using his

example with the appropriate adjustments, the probabilities of matriculation and enrolment can be written as follows:

Stage 1:

Stage 2:

where  $\sigma$  stands for the logisitic distribution. The coefficients  $\beta_1$  are estimated using the entire sample of YAs who have left high school by the end of 2005, while  $\beta_2$  are estimated using the subsample of YAs who have matriculated.  $X$  represents the individual and family variables which are presumed to be relevant in the process of educational attainment.

Bayes' theorem can then be applied to find the unconditional probability of enrolment:

This is derived from the theorem in its original form:

Applied to this example, where  $P(M) = \text{Matric}$ ,  $P(E) = \text{Enrolment}$  and  $P(E|M) = \text{Enrolment given Matric}$  (the probability of matriculation given enrolment is certain as only those who have matriculated are considered in the enrolment decision), and with some rearranging, we obtain

Thus the unconditional probability of enrolment simply involves multiplying together the estimated probabilities from both stages. In this way both the sequential nature of enrolment, as well as the possibility of selection bias, is accounted for.

## 4.2.2 Predicted probability graphs

Predicted probability graphs comprise the second way in which probabilities are used to explore the role of income and ability in educational attainment. While the discrete change in probability assesses the responsiveness of each race to a change in income or ability, graphs examine the racial gap in educational attainment across the entire range of these variables. With graphs it is possible to analyse whether a difference in responsiveness across races results in significantly different predicted probabilities of educational attainment.



In order to test the equality of predicted probabilities (which is done over all included values of income and ability), the delta method is used to calculate the variance of the predicted probability functions (see Long (2009) and Xu and Long (2005) for a derivation of this method). The hypothesis when testing the equality of predicted probabilities is

where the values of the  $x$ s within can be identical or different across groups. This is tested using the following z-statistic

---

Before turning to the analysis, it should be noted that while predicted probabilities have the advantage of being unaffected by residual variation, they are not entirely problem-free. The calculation necessarily includes all of the coefficients in the regression, some of which may be insignificant and therefore should not actually have any bearing on the outcome. If the insignificant coefficients are small, this is not a major issue. However, when the coefficients are large (in absolute terms) but insignificant, they may have a substantial impact on the level of the predicted probabilities which is, in fact, spurious. Determining the extent to which this issue is significant in this analysis is unfortunately beyond the scope of the paper. However, it should be borne in mind when viewing the results.

In the following chapter, the role of income and ability will be estimated using marginal effects, discrete changes in predicted probabilities and predicted probability graphs. The graphs will summarise the findings of the marginal effects and discrete changes in probability, and illustrate the racial gap in educational attainment over the entire range of the variables of interest. Finally, the interaction of income and ability will be considered using just predicted probability graphs.

## 5. ANALYSIS

The findings of Chapters 2 and 3 make it clear that Africans and Coloureds in the Cape Town metropolitan area are not on an equal footing with regards to many socio-economic factors. On average, Africans come from homes with fewer financial resources, lower parental education and greater parent absenteeism. They have more siblings, they attend schools of lower quality and they score substantially and significantly lower on literacy and numeracy measures.

Yet, in spite of these differences, the proportion of Coloureds matriculating is only 8 percent greater than that of Africans. Furthermore, for those who pass Grade 12 and are eligible to pursue higher education, the proportion of Africans and Coloureds who enrol does not differ substantially. This could suggest that different processes are occurring in the educational attainment of each race group. Labour market conditions which seem to favour Coloureds lend support to this suggestion.

The analysis presented in this chapter seeks to shed light on the processes of educational attainment occurring within the African and Coloured YA populations in the Western Cape. The particular focus is on the role of current income and manifest ability. The extent to which these variables determine educational attainment, and how the races compare were they to possess the same individual and background characteristics, are issues which are addressed here.

### 5.1 Income

In this section, marginal effects and discrete changes in probability are presented by race. These estimates are reflections of the sensitivity of each race's educational attainment to a change in income or ability. The discrete change in probability is also used to analyse the *unconditional* probability of enrolment. Thereafter, graphs of the predicted probability of matriculation and enrolment are presented, which allow a racial comparison of educational attainment over the entire income distribution.

#### 5.1.1 Marginal effects

Three regressions are used to calculate the marginal effect of income in this section. The first regression is the simplest, incorporating only individual characteristics and income. The second regression adds the LNE score to the estimation, while the third regression incorporates family factors. As stated already, the progressive addition of variables is based on Carneiro and Heckman (2002), who argue that it is necessary to account for both ability and family background in order to

strip away the effects of long-term income from current income. As variables are added, one would expect the marginal effect of income to decrease in magnitude as the effect of long-term income is increasingly captured by these variables. The extent of the decrease, and whether the coefficient retains significance, is an empirical question which will be answered here. Although all marginal effects are displayed, the focus is largely on income.

## Stage 1 – Matric

**Table 5.1.1: Marginal effects for Africans, Stage 1 (Matric)**

	1	2	3
<b>Age in 2005</b>	-0.870*** (0.243)	-0.828*** (0.245)	-0.875*** (0.248)
<b>Age in 2005 squared</b>	0.017*** (0.006)	0.016*** (0.006)	0.017*** (0.006)
<b>Male (d)</b>	0.106** (0.050)	0.093* (0.051)	0.086* (0.052)
<b>Grade 9 in 2002 (d)</b>	-0.516*** (0.025)	-0.494*** (0.028)	-0.501*** (0.028)
<b>Grade 10 in 2002 (d)</b>	-0.404*** (0.039)	-0.380*** (0.043)	-0.380*** (0.043)
<b>Grade 11 in 2002 (d)</b>	-0.186*** (0.063)	-0.158** (0.066)	-0.167** (0.067)
<b>Income</b>	0.079** (0.031)	0.069** (0.031)	0.052 (0.032)
<b>LNE score</b>		0.117*** (0.029)	0.112*** (0.030)
<b>Mother's education</b>			0.000 (0.010)
<b>Father's education</b>			0.009 (0.011)
<b>Proportion life lived with mother</b>			0.064 (0.096)
<b>Proportion life lived with father</b>			0.076 (0.076)
<b>Siblings</b>			-0.032** (0.016)
<b>Mother's education missing (d)</b>			0.080 (0.122)
<b>Father's education missing (d)</b>			-0.022 (0.107)
<b>N</b>	<b>616</b>	<b>616</b>	<b>616</b>

Notes: 1. The data are weighted to account for sample design and individual non-response. 2. Robust standard errors are in parentheses. 3. (d) for discrete change of dummy variable from 0 to 1. 4. Marginal effects that are significant at 10% are marked with \*, significant at 5% \*\* and significant at 1% \*\*\*.

Table 5.1.1 presents the marginal effects for Africans at the matriculation stage. In Regression 1 it is estimated that a 10 percent increase in per capita monthly income increases the probability of matriculation by 0.79 percent, and this finding is significant at 5 percent. Regression 2 includes the LNE score in the estimation. As indicated by Carneiro and Heckman (2002), the inclusion of manifest ability reduces the coefficient on income because the effect of long-term income is now picked up, at least partially, by this variable. The reduction, however, is relatively small. A 10 percent increase in income is now estimated to raise the probability of matriculation by 0.69 percent. The estimate remains significant at 5 percent.

Regression 3 incorporates family variables into the model. Specifically, parental education, parental co-residence until the age of 14 and the number of siblings are included. There are also two dummy variables to capture missing parental education<sup>12</sup>. The coefficients are small, and none are estimated to be significant except for siblings. However, their inclusion decreases the coefficient of income, in line with the results of Carneiro and Heckman (2002). A 10 percent rise in income is now associated with a 0.52 percent increase in the probability of matriculation and the coefficient is no longer significant at any of the conventional levels.

Table 5.1.2 repeats the previous table, with regards to Coloureds. The estimates are consistently larger than those for Africans, at 2.11, 1.86 and 1.52 percent for Regressions 1, 2 and 3 respectively (in response to a 10 percent change in income). In addition, all estimates are significant at 1 percent. While the pattern of decreasing marginal effects is the same across races, income remains significant once ability and family background have been controlled for among Coloureds but not Africans.

The results for Africans suggest that the relevance of income to matriculation is based on long-term household resources as opposed to short-term credit constraints. In contrast, both current and long-term income may be important determinants of matriculation for Coloureds. Given that many non-White children grow up in poor areas within resource-constrained households, but may attend school for little to no fees, the lack of importance of current income for Africans is not altogether surprising. However, the fact that it is relevant for Coloureds, who have the higher household per capita income on average, is unexpected. Rather, one would expect that should credit constraints be present, they would be more binding for the group which is poorer on average.

---

<sup>12</sup> Those with missing data have been assigned a value of one for the dummy variable and zero for the continuous variable in order to keep them in the sample and avoid issues related to “selective sample composition due to non-random missing information” (Lauer, 2003:238).

**Table 5.1.2: Marginal effects for Coloureds, Stage 1 (Matric)**

	1	2	3
<b>Age in 2005</b>	-0.570 (0.434)	-0.270 (0.450)	-0.287 (0.462)
<b>Age in 2005 squared</b>	0.009 (0.011)	0.002 (0.011)	0.002 (0.012)
<b>Male (d)</b>	0.004 (0.052)	-0.049 (0.055)	-0.042 (0.057)
<b>Grade 9 in 2002 (d)</b>	-0.682*** (0.033)	-0.646*** (0.039)	-0.649*** (0.039)
<b>Grade 10 in 2002 (d)</b>	-0.608*** (0.040)	-0.569*** (0.047)	-0.575*** (0.048)
<b>Grade 11 in 2002 (d)</b>	-0.339*** (0.067)	-0.309*** (0.072)	-0.307*** (0.074)
<b>Income</b>	0.211*** (0.034)	0.186*** (0.034)	0.152*** (0.038)
<b>LNE score</b>		0.200*** (0.036)	0.190*** (0.038)
<b>Mother's education</b>			0.010 (0.012)
<b>Father's education</b>			0.024** (0.011)
<b>Proportion life lived with mother</b>			0.079 (0.132)
<b>Proportion life lived with father</b>			-0.064 (0.088)
<b>Siblings</b>			0.049** (0.022)
<b>Mother's education missing (d)</b>			-0.105 (0.147)
<b>Father's education missing (d)</b>			0.165 (0.120)
<b>N</b>	<b>673</b>	<b>673</b>	<b>673</b>

Notes: 1. The data are weighted to account for sample design and individual non-response. 2. Robust standard errors are in parentheses. 3. (d) for discrete change of dummy variable from 0 to 1. 4. Marginal effects that are significant at 10% are marked with \*, significant at 5% \*\* and significant at 1% \*\*\*.

As the final section of the literature review showed, experience in the labour market differs substantially by race, with Coloureds faring better on almost every measure. For example, in the two years after leaving school, Coloureds without matric had greater success in finding employment than Africans *with* matric (Lam et al 2009). This has a bearing on the opportunity cost of remaining in school, which is determined by the alternative activities that a YA could be doing during this time. If it is possible to enter the labour market, the opportunity cost is the expected wage, which rises with the probability of employment. This implies that the opportunity cost to remaining in school is higher for Coloureds than Africans, as they are more likely to find work.

If it is assumed that YAs weigh up the costs and benefits of attaining education in line with standard human capital theory regarding schooling decisions (Lam et al 2010), the higher opportunity cost faced by Coloureds may explain their greater responsiveness to an increase in income. Africans, on the other hand, may choose to remain in school whether or not there is an increase in household income, as the opportunity cost of this choice is essentially negligible.

Note that for both races, most of the family variables are found to be insignificant in the previous regressions. Given that these variables tend to be collinear to varying degrees, this could be due to difficulty in separating out the individual effects. A Wald test is used to determine if the variables are jointly significant despite being individually insignificant. This test is based on a  $\chi^2$  distribution, with the null hypothesis that the included coefficients are jointly insignificant. The null hypothesis cannot be rejected at any conventional level of significance for the seven family variables of Africans, whereas it is rejected at 10 percent for Coloureds (results not shown). Focusing just on parental education, the same conclusion is drawn. This seems to suggest that family background is not relevant to the process of educational attainment at the secondary level for Africans, while it is for Coloureds. While it should be borne in mind that this is not what is found in the international literature, these variables are included as controls and further exploration of their insignificance is not included here.

## **Stage 2 – Enrolment into post-secondary education**

The regressions for Stage 2 repeat those of Stage 1 with one exception. Rather than including the grade of enrolment in 2002, all regressions include the year in which the YA successfully matriculated. Note that the regressions calculate the marginal effects on the probability of enrolment *given* matriculation. This means that the estimation sample is restricted to those in the full sample who passed matric between 2002 and 2005. As a result, the sample size drops by more than half for each race group. When interpreting the marginal effect, it must be viewed as the additional impact of a variable given the role it has already played in determining matriculation. The *unconditional* probability of enrolment will be dealt with in the following section of the analysis.

**Table 5.1.3: Marginal effects for Africans, Stage 2 (Enrolment in PSE)**

	<b>1</b>	<b>2</b>	<b>3</b>
<b>Age in 2005</b>	-0.010 (0.347)	0.043 (0.359)	0.038 (0.362)
<b>Age in 2005 squared</b>	-0.002 (0.008)	-0.003 (0.009)	-0.003 (0.009)
<b>Male (d)</b>	-0.017 (0.072)	-0.030 (0.073)	-0.028 (0.076)
<b>Matriculated in 2002 (d)</b>	0.384*** (0.076)	0.350*** (0.086)	0.350*** (0.088)
<b>Matriculated in 2003 (d)</b>	0.210** (0.101)	0.176* (0.106)	0.183* (0.111)
<b>Matriculated in 2004 (d)</b>	0.013 (0.108)	-0.011 (0.110)	0.002 (0.115)
<b>Income</b>	0.116*** (0.044)	0.102** (0.044)	0.037 (0.050)
<b>LNE score</b>		0.087** (0.042)	0.084* (0.044)
<b>Mother's education</b>			0.030* (0.016)
<b>Father's education</b>			0.020 (0.016)
<b>Proportion life lived with mother</b>			-0.066 (0.140)
<b>Proportion life lived with father</b>			0.148 (0.104)
<b>Siblings</b>			-0.027 (0.022)
<b>Mother's education missing (d)</b>			0.266* (0.157)
<b>Father's education missing (d)</b>			0.186 (0.160)
<b>N</b>	<b>260</b>	<b>260</b>	<b>260</b>

Notes: 1. The data are weighted to account for sample design and individual non-response. 2. Robust standard errors are in parentheses. 3. (d) for discrete change of dummy variable from 0 to 1. 4. Marginal effects that are significant at 10% are marked with \*, significant at 5% \*\* and significant at 1% \*\*\*.

Regression 1 in Table 5.1.3 estimates a 1.16 percent increase in the probability of enrolment in response to a 10 percent increase in income, ceteris paribus. Adding the LNE score to the second regression lowers the size of the coefficient to 1.02 percent and reduces the significance to 5 percent. The inclusion of family variables in Regression 3 causes a substantial drop in the coefficient, to 0.37 percent, and it becomes insignificant at all conventional levels. As was the case at the matriculation stage, income does not appear to have significant individual explanatory power among Africans once ability and family background factors have been accounted for.

Turning to Table 5.1.4 which looks at Coloureds, the probability of enrolment rises by 2.75, 2.53 and 1.57 percent in Regressions 1, 2 and 3 respectively in response to a 10 percent increase in income. All estimates are significant at one percent. In line with the matric stage, income retains individual explanatory power once all other background factors have been accounted for, but its marginal effect decreases with each progressive addition.

**Table 5.1.4: Marginal effects for Coloureds, Stage 2 (Enrolment in PSE)**

	1	2	3
<b>Age in 2005</b>	1.045*	1.296**	1.159*
	(0.602)	(0.632)	(0.645)
<b>Age in 2005 squared</b>	-0.026*	-0.032**	-0.029*
	(0.015)	(0.015)	(0.016)
<b>Male (d)</b>	0.106*	0.071	0.100
	(0.063)	(0.065)	(0.067)
<b>Matriculated in 2002 (d)</b>	0.108	-0.015	0.012
	(0.179)	(0.191)	(0.195)
<b>Matriculated in 2003 (d)</b>	0.034	-0.047	-0.031
	(0.164)	(0.168)	(0.173)
<b>Matriculated in 2004 (d)</b>	-0.078	-0.135	-0.124
	(0.131)	(0.133)	(0.135)
<b>Income</b>	0.275***	0.253***	0.157***
	(0.052)	(0.051)	(0.058)
<b>LNE score</b>		0.133***	0.106**
		(0.046)	(0.047)
<b>Mother's education</b>			0.046***
			(0.015)
<b>Father's education</b>			0.006
			(0.015)
<b>Proportion life lived with mother</b>			0.106
			(0.201)
<b>Proportion life lived with father</b>			0.105
			(0.116)
<b>Siblings</b>			-0.050
			(0.032)
<b>Mother's education missing (d)</b>			0.031
			(0.287)
<b>Father's education missing (d)</b>			0.093
			(0.179)
<b>N</b>	<b>326</b>	<b>326</b>	<b>326</b>

Notes: 1. The data are weighted to account for sample design and individual non-response. 2. Robust standard errors are in parentheses. 3. (d) for discrete change of dummy variable from 0 to 1. 4. Marginal effects that are significant at 10% are marked with \*, significant at 5% \*\* and significant at 1% \*\*\*.

It appears that the conclusion regarding the role of income in educational attainment is the same across stages: current income is relevant for Coloureds but not Africans in both matriculation and



enrolment. Rather, for Africans, the role of income seems to derive solely from the relationship between long-term income and educational attainment.

It is plausible that the argument regarding the role of differential labour market conditions in matriculation also applies to enrolment in PSE: if Coloureds face a greater opportunity cost to remaining in the education system, the availability of household income will play a larger role in their decision to enrol. Turning again to Lam et al (2009), it is reported that among Cape Town YAs with matric, Coloureds find employment much more quickly than Africans after leaving school. The gap between the races is staggering. For example, roughly 70 percent of Coloured men are working 6 months after completing Grade 12 as opposed to less than 20 percent of African men. After two years, the gap remains substantial at approximately 30 percent. With regards to women, the gap is smaller but still considerable. It would be highly useful to compare the employment outcomes of Africans with some form of tertiary education to Coloureds with just matric. Unfortunately no such analyses were found in the available literature. If the racial imbalance in employment outcomes persists even when Africans hold some form of PSE, this would serve as further evidence that the opportunity cost to tertiary education is substantially higher for Coloureds than Africans.

Fortunately a comparison of the average hourly wage at different levels of education is possible. Data from the 2006 LFS show that among the wage-employed with matric, Coloureds earn almost 60 percent more than Africans, and that earnings are only equalised at the post-secondary level (Hofmeyr 2008). This is additional evidence that the opportunity cost of tertiary education is higher for Coloureds than Africans. Moreover, as the hourly wage of Africans essentially triples once PSE is obtained, as opposed to increasing by slightly less than double for Coloureds (Hofmeyr 2008), the incentive to pursue further studies is substantially greater for Africans.

The fact that Africans face more limited employment opportunities with matric, coupled with dramatically improved labour market outcomes with tertiary education, may explain why current income does not feature significantly in their enrolment decision while it does for Coloureds. Simply put, the return to pursuing further education for an eligible African youth is so high in relation to its opportunity cost that enrolling makes sense no matter how resource constrained the household is. However, direct costs must also be borne in mind. At the secondary level, direct costs were relatively unimportant as fees at government schools are generally quite low, and may be waived for students who cannot afford them. In contrast, many tertiary institutions charge comparatively high fees and are not legally obligated to accept students who cannot afford to pay.

Table 5.1.5 uses expenditure data in CAPS to compare mean secondary and post-secondary fees in 2002, 2005 and 2006 for the sample used in this analysis<sup>13</sup>. These figures have been adjusted for inflation and are reported in 2008 prices. The table makes it clear that tertiary education is substantially more expensive than secondary school. In light of the fact that median annual household income for Africans was R17 500 in 2005 (Lam et al 2010: 6), or R21875 in 2008 prices, it is presumed that a student loan must be necessary for members of poorer households who enrol in PSE. However, it is precisely these households which are in the weakest position to borrow as they generally possess little to no collateral. As a result, it is likely that paying for PSE is a severe burden for many African students (Lam et al 2010), and may prevent them from enrolling.

**Table 5.1.5: Average secondary and post-secondary fees**

	Mean	95% C.I. lower bound	95% C.I. upper bound
<b>Secondary school fees - 2002</b> (n=1140)	1313.66 (119.72)	1078.76	1548.56
<b>Post-secondary fees - 2005</b> (n=134)	11119.9 (1004.83)	9132.38	13107.41
<b>Post-secondary fees - 2006</b> (n=115)	15709.17 (985.81)	13756.30	17662.05

Notes: 1. The data are weighted to account for sample design and individual non-response. 2. Standard errors are in parentheses. 3. All values are reported in 2008 prices, using the CPI deflator of 72, 80 and 83.7 respectively (CPI deflators obtained from Statistics South Africa).

In view of the preceding discussion, one would expect current income to be significant in the enrolment decision of Africans. Taken at face value, its insignificance in the regressions implies the existence of perfectly functioning credit markets. In such a situation, Africans who are eligible to enrol in PSE can and do obtain loans to finance their studies, regardless of their household income. However, we know that this is not the case in South Africa, making the apparent irrelevance of current income a conundrum. It may be that African YAs are accessing funds from alternative sources (such as family, friends or micro-financing agencies), which will provide loans in the absence of sufficient collateral. In the case of micro-financing agencies, these loans can be exorbitantly expensive in order to compensate for the riskiness of the debt. However, in light of the large employment and wage returns to PSE, Africans may nonetheless turn to this source of credit if they are left with no other option.

<sup>13</sup> Information from both 2005 and 2006 is used to calculate average post-secondary fees in order to compensate for the small number of observations available in each of the years.

### 5.1.2 Discrete changes in probability

The preceding section presented the marginal effects of all included variables, focusing specifically on income, and revealed that current income is significant for Coloureds but not for Africans, once all factors are accounted for. This section looks at the effect of a standard deviation change in income on the probability of matriculation and enrolment, holding all other included variables at a chosen set of values. The aim of this section is three-fold. First, these figures allow the effect of a larger change in income to be calculated. The results will not paint a different story to the marginal effects, but will provide a clearer idea of the responsiveness of educational attainment to income. The estimates may also be more accurate if the change occurs over a section of the probability curve which is non-linear (Long 2006). Second, using a standard deviation change allows the effects of income and ability to be compared. These variables are measured in different units and cannot be compared directly. Third, this methodology allows the effect of a change in income on the *unconditional* probability of enrolment to be calculated.

With both marginal effect and probability calculations, it is necessary to choose values for all of the included covariates. This is a restriction which is not present in linear models, and it means that conclusions cannot be widely generalised. In the previous section which dealt with marginal effects, all variables were set to the point of means<sup>14</sup>. Here a set of characteristics has been chosen with the following characteristics: the YA is male, enrolled in Grade 11 in 2002, and has no missing information for parental education. In the matric subsample, the year of matriculation is set to 2003. In both stages, the remaining continuous variables are set to the aggregate mean of the full sample and the matric subsample<sup>15</sup>. Note that Africans and Coloureds are assigned exactly the same characteristics and it is therefore the coefficients that drive any difference that is found between the races.

---

<sup>14</sup> A shortcoming of this method is that discrete covariates (such as gender) are also set to the mean, which is equal to the proportion of the sample with a value of 1 for that variable. This is clearly an impossible value for a discrete variable to take in reality. For example, no one is 0.49 male. They are either male, and take on a value of 1, or female, with a value of 0.

<sup>15</sup> The only exception is age. In the first stage, age is set to the mean of all YAs in Grade 11 in 2002, while in the second stage it is set to the mean of all YAs who matriculated in 2003. This is done in order to ensure that the characteristics are as realistic as possible.

## Matriculation and conditional enrolment

Table 5.1.6 presents the change in the probability of matriculation and conditional enrolment in response to a standard deviation change in income<sup>16</sup>. Regressions 1 – 3 refer to the same regressions which were discussed above in the marginal effects section.

**Table 5.1.6: Discrete change in the probability of matric and conditional enrolment for a standard deviation change in income around the mean**

	Matric		PSE   Matric	
	Africans	Coloureds	Africans	Coloureds
<b>Regression 1</b>	0.048	0.196	0.112	0.266
<b>Regression 2</b>	0.039	0.181	0.099	0.233
<b>Regression 3</b>	0.027	0.149	0.037	0.135

**Note:** The levels of the probabilities from which the changes are calculated are reported in Tables A.1 and A.2 in the Appendix.

Table 5.1.6 summarises the findings from the marginal effects section. Coloured youths in the Western Cape are substantially more responsive to a change in income at both stages of educational attainment. Specifically, Coloureds are two to five times more responsive to a standard deviation change in income in all regressions. With regards to each race's responsiveness to current income across stages, Regression 3 indicates that a standard deviation increase in this variable has approximately the same effect at both matric and enrolment.

It should be borne in mind that the probability of an African youth matriculating at the initial, lower level of income is substantially higher than the equivalent probability for a Coloured youth in all three regressions (see Table A.1 and A.2 in the Appendix). For example, using the figures estimated by Regression 3, the probability of an African YA matriculating is 0.74 at the lower level of income, as opposed to only 0.32 for an otherwise identical Coloured youth<sup>17</sup>. This may explain in part why a change in income has a smaller effect on Africans than Coloureds: the majority of Africans are

<sup>16</sup> The changes in probability are obtained by calculating the probability of the event (matriculation or enrolment) given Set A of characteristics, calculating the probability of the event given Set B of characteristics, and subtracting the first result from the second (Long & Freese 2006:171). In Set A, income is held at half a standard deviation below the mean, while in Set B, income is held at half a standard deviation above the mean. All other characteristics are held constant. The mean referred to is the average income over both Africans and Coloureds. At the matric stage, it is calculated using the full sample (1289 observations), while at the enrolment stage, the subsample of those who passed matric is used (586 observations).

<sup>17</sup> The predictions seem at odds with the findings in Table 4 of Chapter 3, where it was reported that 42 percent of Africans and 50 percent of Coloureds matriculated between 2002 and 2005. This issue is dealt with in more detail in the section on predicted probability graphs. Suffice it to say at this point that the disparity in between the empirical rate and the predicted probability arises in part from the values at which the races are held, particularly age.

already predicted to matriculate at the lower income, indicating that an increase in income is likely to have a smaller effect on the matriculation rate. Note, however, that this argument cannot be used to explain the greater responsiveness of Coloureds at the enrolment stage. Although Africans are predicted to have a higher enrolment rate than Coloureds at the lower income, their probability is below 50 percent, indicating that there is still room for substantial improvement.

It may seem that some of the calculated changes in probability shown in Table 5.1.5 are unrealistically large given the marginal effects discussed in the previous section. However, their magnitude makes sense when the mean and standard deviation of income are considered. Referring to the actual figures (rather than their logged and normalised counterparts), mean weighted per capita income for the full sample is R841.12, with a standard deviation of R897.50. A change in income of one standard deviation around the mean therefore represents a substantial improvement in financial well-being from R392.37 to R1289.95. Given this, the larger changes in probability seem reasonable. In fact, what is perhaps unexpected is how small some of the changes in probability are for the African subsample. The fact that the likelihood of an African YA matriculating or enrolling remains essentially unchanged in response to such a large increase in income emphasises the finding that current income has little to no relevance in the process of educational attainment for Africans.

## **Enrolment – conditional versus unconditional estimates**

As outlined in the previous chapter, matriculation and enrolment in South Africa are generally viewed as sequential events because the completion of Grade 12 is required by many tertiary institutions. This must be taken into account when analysing the determinants of enrolment. However, those who pass matric are a select sample of the full sample, and as such the conditional estimates may suffer from selection bias. If sample selection bias is a problem, one would expect to find that should non-matriculants be able to enrol, an increase in income may affect them differently to those who do matriculate.

The following table seeks to answer this question. The first pair of columns presents the change in the probability of enrolment conditional on matriculation (copied unaltered from Table 5.1.6). The remaining columns present the discrete change in the probability of unconditional enrolment using two methods which were outlined in the previous chapter. In brief, the first method includes the full sample of YAs attending secondary school in the enrolment regression. The second method uses Bayes' Law, and involves multiplying the predicted probability of matriculation by the predicted probability of enrolment. This gives the unconditional probability of enrolment, and the change in probability is calculated thereafter.

**Table 5.1.7: The change in the probability of enrolment for a standard deviation change in income around the mean**

	PSE   Matric		PSE: Single Regression		PSE: Bayes' Law	
	Africans	Coloureds	Africans	Coloureds	Africans	Coloureds
<b>Regression 1</b>	0.112	0.266	0.145	0.195	0.120	0.245
<b>Regression 2</b>	0.099	0.233	0.128	0.137	0.105	0.192
<b>Regression 3</b>	0.037	0.135	0.076	0.075	0.045	0.118

**Note:** The levels of the probabilities from which the changes are calculated are reported in Tables A.2 – A.4 in the Appendix.

The conditional enrolment estimates could suffer from selection bias, as stated above. On the other hand, using a single regression ignores the fact that matriculation and enrolment are, by and large, sequential events and assumes that anyone who has been enrolled in high school (until at least Grade 9) may pursue further studies. The estimates created with the use of Bayes' Law represent a compromise between these approaches. They incorporate the sequential nature of educational attainment while also accounting for the fact that the probability of matriculation is not equal across individuals.

In comparison to the conditional estimates, the discrete changes in probability calculated with the single regression are consistently larger for Africans and smaller for Coloureds across Regression 1 - 3. This indicates that income is estimated to have a greater effect on enrolment when the process of educational attainment is treated sequentially for Coloureds, while the reverse is true for Africans. The estimates using Bayes' Law fall between the conditional and single regression figures. On the whole, they appear to be slightly closer to the estimates generated in the conditional regression than those from the single regression.

The fact that the Bayes' Law estimates are not identical to those from the single regression implies that it is necessary to account for the sequential nature of educational attainment when assessing the impact that factors such as income have on enrolment. Moreover, as they are also not identical to the conditional estimates, sample selection bias may be an issue. However, the difference is not dramatic, suggesting that this bias may not be too severe.

### 5.1.3 Predicted probability graphs

This section of the analysis presents graphs of the predicted probability of matriculation and enrolment over the entire range of income. The question here is not whether the responsiveness to income varies by race, but rather how the responsiveness translates into differences in the predicted probability of educational attainment between Africans and Coloureds over the entire range of income.

In the descriptive statistics presented in Table 3.4 of Chapter 3, it was found that Coloureds had a significantly higher matriculation rate than Africans (50 versus 42 percent), while the rate of enrolment (given matriculation) was not statistically different across the two groups (48 versus 46 percent). Table 3.4 also demonstrated that Africans fare more poorly than Coloureds on many characteristics which are correlated with educational attainment, such as income, LNE score, parental education and parental co-residence. Predicted probabilities are able to account for these factors and hold them at specific values. This allows an examination of the predicted probability of educational attainment under different hypothetical scenarios.

Three scenarios are depicted in graphical form. Firstly, how do the races compare over the income range when they are held to their own average characteristics? Secondly, how do the races compare when they are held equal on all characteristics except for age? Thirdly and finally, how do the races compare when they are equal in every way?

The second scenario, where the races are equated except for age, is included because of the realities of the education system in South Africa. Referring to the estimation sample, the average age of Africans enrolled in Grade 11 in 2002 is 18.59 years old as opposed to 16.97 for Coloureds. This is partly a result of a higher failure rate among Africans; on average, African YAs had failed 0.93 grades by Wave 1 as opposed to 0.56 grades for Coloureds<sup>18</sup>. Lam et al (2008) argue that poor quality schooling, especially weak evaluation, leads to high rates of grade failure and repetition. Thus the older average age of Africans in part reflects the inferior quality of the schooling they receive. As such, it is interesting to compare educational outcomes by race when individual and home characteristics are equal but the quality of schooling, proxied (weakly) by age, is not.

The graphs are based on estimates from Regression 3, implying that the income variable is assumed to capture only current income. Each graph is comprised of two curves, representing the probability of matriculation (and then enrolment) for Africans and Coloureds. Where the lines are solid, the difference in the predicated probabilities of the two races is estimated to be significant at 5 percent. The dashed portions of the lines represent an insignificant difference between races.

The idea of common support which was discussed in the descriptive statistics chapter is relevant here. In each graph there are dashed lines indicating the points on the x axis where logged, normalised income is equal to -1 and 1. The included income range represents the area of common support for the two races. While just over 25 percent of the African sample has an income below -1, only slightly more than 4 percent of the Coloured sample falls in this income range. The situation is

---

<sup>18</sup> It is also a result of Africans starting school at an older age. For example, among YAs enrolled in Grade 11 in 2002 who had *not* failed any grades, Africans were on average 1.17 years older than Coloureds.

reversed at the upper end of the distribution: only 2.4 percent of the African sample has an income greater than 1 as compared to approximately 25 percent of Coloureds. While statistical software can easily generate predicted probabilities in regions of the data space which are relatively sparsely populated, these predictions must be viewed with caution. In the following discussion, only the area of common support will be used to compare the educational attainment across Africans and Coloureds.

Note that the same hypothetical individual is considered in this section as in the discrete change calculations. Thus the predicted probabilities refer to a male who is in Grade 11 in 2002, matriculates in 2003, and has no missing information.

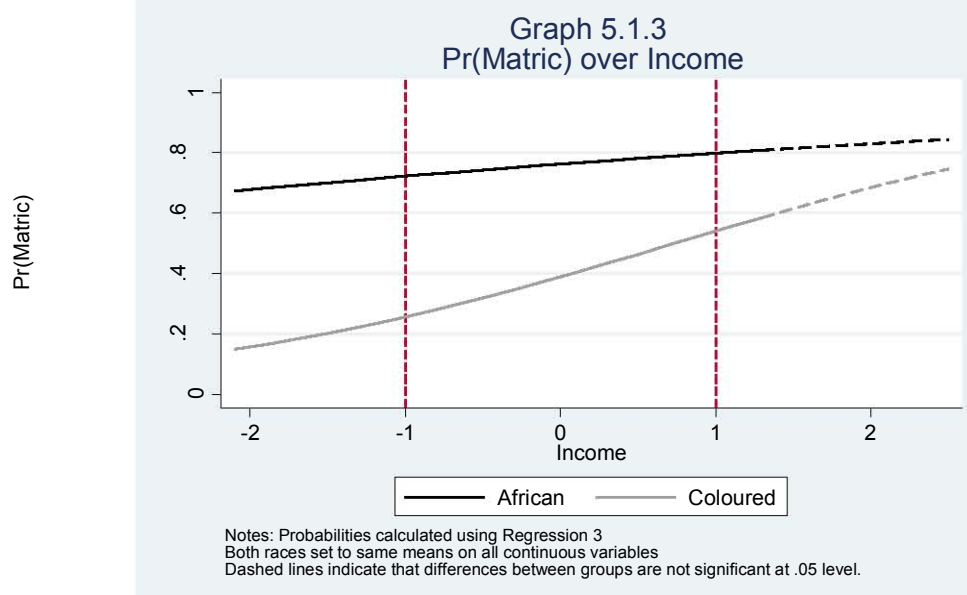
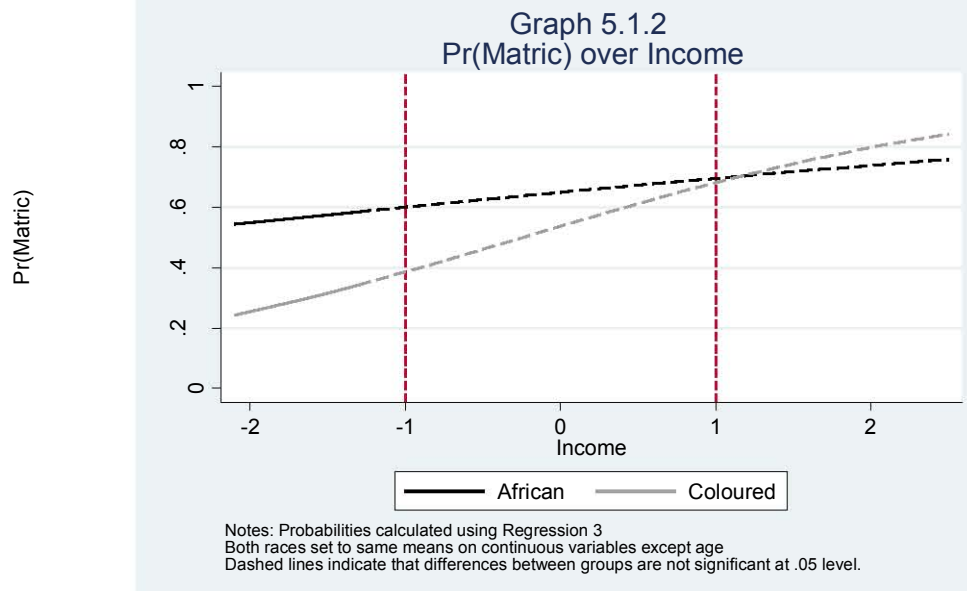
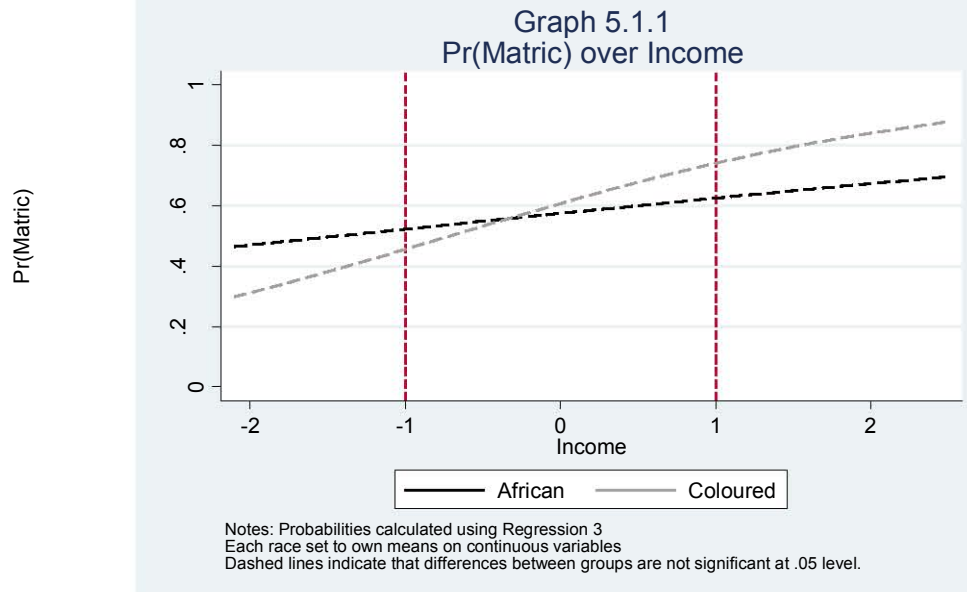
### **Stage 1 - Matric**

Graph 5.1.1 reveals that the probability of matriculation does not differ significantly between Africans and Coloureds at any level of income when each race is held to their own average characteristics (see Table A.5 in the Appendix for the values of these averages). This may suggest that the higher empirical rate of matriculation among Coloureds is due to the gap in income between the races. This makes sense light of the lowess estimate in Figure 1 of Chapter 3, which shows that the proportion of Coloureds matriculating is relatively high in the upper tail of the income distribution where no Africans are found.

The next graph equates the races on all bases except age. As a result, Africans are predicted to have a higher matriculation rate over most of values of income. However, the gap is not significant over the range  $[-1;1]$ . The third graph shows that when the races are also equated on age, the gap increases substantially. Within the area of common support, Africans are found to be between 20 – 40 percent more likely to matriculate than Coloureds (significant at 5 percent).

This set of graphs suggests several things. Firstly, equalising current income across Africans and Coloureds removes the small advantage which Coloureds have in the probability of matriculation. Secondly, equalising the races on all bases except age improves the outcomes of Africans in relation to Coloureds, but not to a statistically significant degree. It is the equalisation of age in addition to all other characteristics which makes Africans significantly and substantially more likely to matriculate than Coloureds. To the degree that age is a proxy for school quality, the fact that this large and significant gap only appears when age is held constant might imply that the education system is a major component of Africans' poor educational outcomes. However, as age is assumed to be at most a weak proxy for school quality, this suggestion is not emphasized.





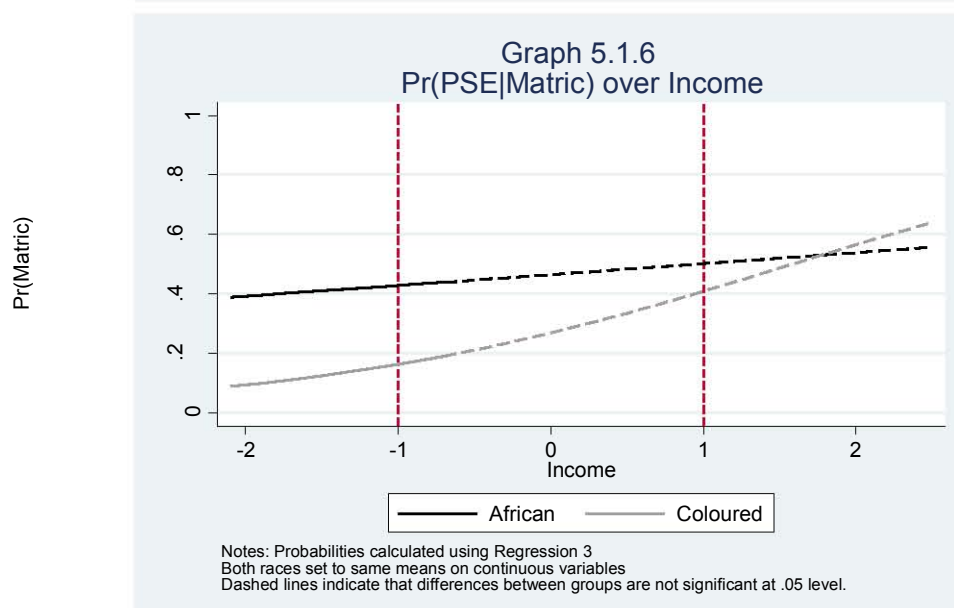
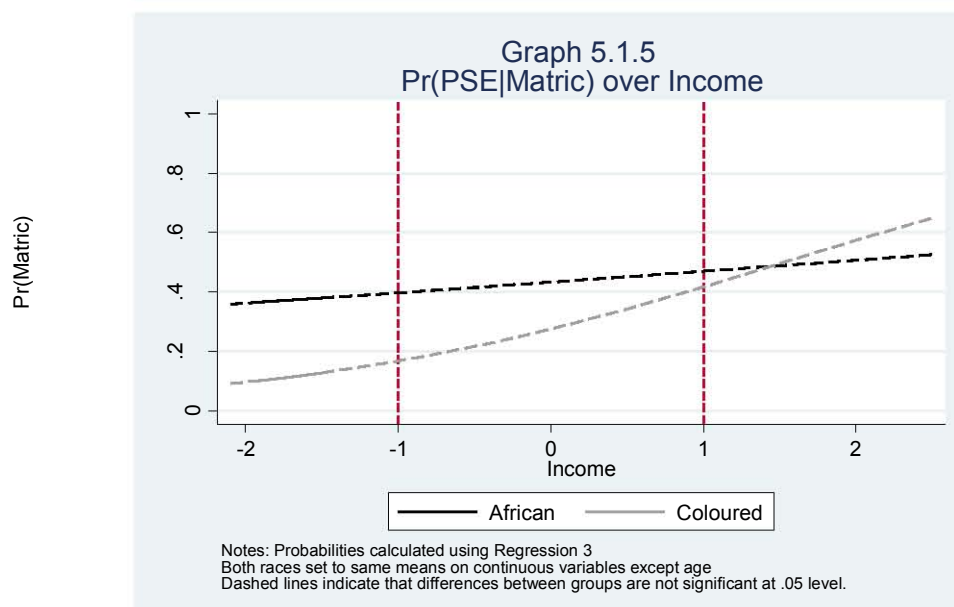
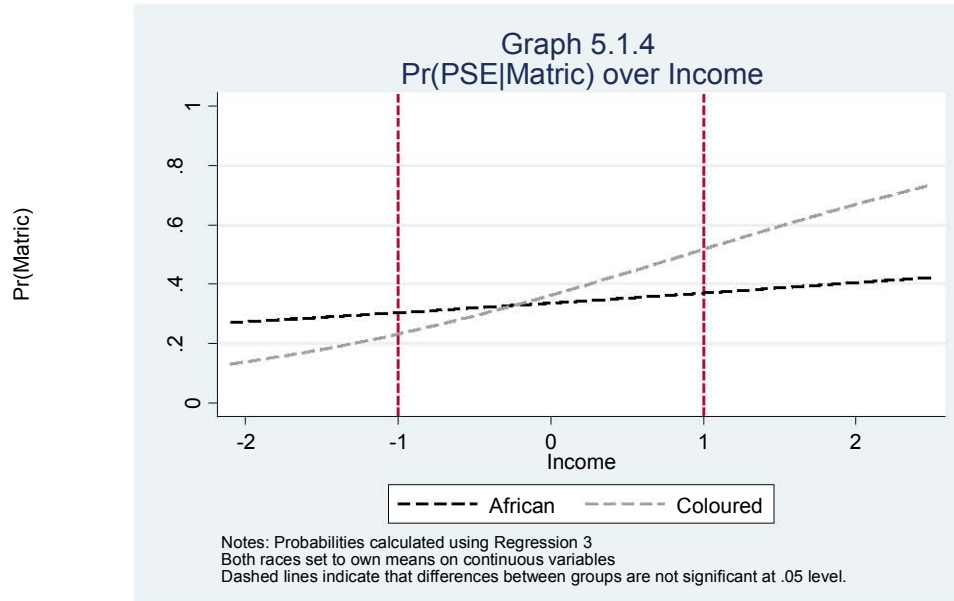
Note, however, that as labour market conditions are not accounted for in this model, the opportunity cost of staying in school has not been controlled for in any of the predictions. The third graph therefore assumes that Africans and Coloureds are equal in terms of age, ability, income and home situation but remain starkly different in terms of the employment and wage opportunities they face. If the hypothesis regarding the role that these opportunities play in educational decision making is correct, one might expect to see a far smaller, if perhaps non-existent, gap between the races should their experience in the labour market be accounted for and held constant.

## **Stage 2 – Enrolment into post-secondary education**

Graphs 5.1.4 – 5.1.6 repeat the previous set of graphs with respect to enrolment in post-secondary education. As the lines are dashed for their entire length in the first graph, the gap between the races is insignificant at all income values when the races are held to their own means. This seems somewhat unexpected in light of the lowess estimate in Figure 3 of Chapter 3, which estimated a greater proportion of Africans enrolling at all income levels between -1 and 1. However, lowess estimates are silent on the matter of significance, and fail to control for any other individual and family characteristics. Indeed, using weighted sample averages, Table 3.4 showed that the enrolment rate did not differ significantly by race. Equalising current income may therefore not cause a significant difference to appear between the races.

Graph 5.1.5 reveals that equalising the races on all bases except age again pushes Africans above Coloureds, but leaves the gap insignificant. Setting age equal across races widens the gap slightly, causing Africans to be significantly more likely to enrol at income values in the lower end of the distribution.

It should be noted that the sample size used in the enrolment regressions is substantially smaller than in the matriculation regressions, including only 260 Africans and 326 Coloureds. While there is no hard and fast rule regarding the minimum sample size for logit regressions, Long and Freese (2006) suggest that less than 100 is inadvisable, and greater than 500 should be sufficient. With fewer observations come wider confidence intervals as point estimates are measured less precisely. This could cause the level of significance to be underestimated in the previous graphs. Whether or not this is the case is an empirical question, and one that unfortunately cannot be answered in this paper given the available data.



Once again the differential labour market conditions need to be borne in mind when interpreting these results. As Coloureds can expect to find work relatively easily with just a matric, the opportunity cost of enrolling is relatively greater at the lower end of the income distribution. This is because the gap between the income these individuals receive at home and their expected wage is larger for those who are poorer, and this is expected to encourage them to enter the labour market. Thus the low rate of enrolment at low income levels is logical given the options available to Coloureds. As household income increases, the difference between what they could be earning and what they receive at home falls, or becomes negative, and accordingly the rate of enrolment rises.

In contrast, Africans face an opportunity cost to enrolment which is relatively low, and consequently an increase in current income does not affect their decision to enrol to a large degree. While their rate of enrolment is significantly higher at the lower end of the distribution, as household income increases the Coloured enrolment rate rises at a faster rate than for Africans, and the gap becomes insignificant.

These graphs raise an interesting question when we remember that they are estimates of the *conditional* enrolment rate. All YAs included in these calculations are matriculants and therefore eligible to enrol in some form of tertiary education. If credit constraints are not preventing Africans from enrolling, and the labour market opportunities for this group strongly favour tertiary education, why are more Africans not enrolling?

The arguments of de Broucker (2005) and more particularly Beattie (2002) may answer this question. Both authors note that YAs from relatively disadvantaged backgrounds may not realise the value of continuing with their education. Beattie (2002) refers to a theory by Wilson (1980, 1987) to explain that YAs from poor households are likely to come from areas where there is high unemployment. This means that they have access to a limited number of role models of people who have found work, especially work that is well-paid or secure. While Beattie is referring to the United States, her statement that “historical discrimination in education and the labour market may prompt black adolescents quite rationally to be sceptical of the link between education and earnings” could just as well have been written with South Africa in mind (2002: 35).

It therefore could be that the relatively low rate of enrolment among Africans, even when assigned the same characteristics as Coloureds, could be due to a disconnect between reality and expectations regarding the benefit provided by further education in the labour market. The lack of significance of income in the enrolment decision of Africans may not reflect the absence of credit constraints. Rather, it could be that many African youth are discouraged from applying to PSE

regardless of their family income. This may be supported by a very recent study by Gurgand et al (2011) who indicate that credit constraints may decrease the enrolment rate into higher education by as much as 20 percentage points among YAs who want to pursue further studies.

## 5.2 Ability

This section follows the same outline as the analysis of income, and its aim is to add depth to the analysis presented thus far. It appears that income affects the process of educational attainment differently for Africans and Coloureds, perhaps in part because of the labour market bias towards Coloureds, as well as a lack of information reaching poorer individuals about the benefits of higher education. Seeing how each race responds to changes in manifest ability may provide a deeper understanding of African and Coloured young adults make their decisions regarding matriculation and enrolment.

### 5.2.1 Marginal effects

In the regressions dealing with income, the first regression included only income and individual characteristics, the second incorporated ability and the third included family characteristics. The addition of variables proceeds in the same order in this section, with the exception that ability is included in all regressions, with income incorporated in the second estimation. As the marginal effects for all variables have been presented in the previous section, only those relating to ability are included here.

When analysing the role of income, the progressive addition of variables separated out the effects of current income from long-term income, in line with Carneiro and Heckman (2002). There is no comparable theory on which to base the addition of income and family background variables when estimating the marginal effect of ability. However, the same procedure is carried out to provide comparability across sections. More importantly, it is necessary to include income and family background in order to ensure that ability is not picking up their effects. This is not an attempt to *remove* the influence of family background and income on the LNE score, as that would result in a measure of innate rather than manifest ability and requires a different technique (e.g. see Branson et al 2009). Rather, it is to ensure that the LNE score reflects the role of acquired skills and knowledge on educational attainment, free of any separate influence which income and the home environment may have on matriculation and enrolment.

## Stage 1 – Matric

With regards to Africans, it is estimated that a unit increase in the LNE score from 0 - 1 (which is close but not exactly equal to a standard deviation increase) results in an African youth being 12.2 percent more likely to matriculate when income and family variables are excluded. Their addition causes the estimate to decline by 1 percentage point, but leaves the significance untouched at 1 percent. Turning to Coloureds, the first estimate is 22.2 percent, falling to 19 percent with the addition of income and family variables. The significance remains unchanged at 1 percent.

**Table 5.2.1: Marginal effects for Africans and Coloureds**

	Stage 1 - Matric		Stage 2 - PSE	
	Africans	Coloureds	Africans	Coloureds
<b>Regression 1</b>	0.122***	0.222***	0.105**	0.168***
<b>Regression 2</b>	0.117***	0.200***	0.087**	0.133***
<b>Regression 3</b>	0.112***	0.190***	0.084*	0.106**

Notes: Significance levels: \* 10%, \*\* 5% and \*\*\* 1 %

Comparing the Coloured figures to those of Africans suggests that ability plays a significant role in matriculation for both races, but that Coloureds are more affected by an increase in ability, just as they are more responsive to an increase in current income. Note that the small decrease in the coefficient on ability following the addition of income and family background implies that the LNE score by and large does not pick up the separate effects that these variables have on educational attainment.

## Stage 2 – Enrolment into post-secondary education

Looking at enrolment for Africans, the table reveals that the role of ability is reduced at the post-secondary stage in comparison to the matriculation stage. The estimates are 10.5, 8.7 and 8.2 percent respectively (in response to a unit change in LNE score) and all are significant. For Coloureds, the estimates decrease from 16.8 to 10.6 percent over the three regressions, and are significant at a level of at least 5 percent. Thus for both races, the impact of ability is lower at the tertiary than secondary level. Furthermore, Coloureds are more sensitive than Africans to a change in ability at both stages.

### 5.2.2 Discrete changes in probability

The discrete changes in the predicted probability of matriculation and enrolment for a standard deviation increase in the LNE score over the three regressions are presented here. As the standard

deviation of this variable is close to one, the estimates presented in the following table are not dramatically different to the marginal effects discussed above. The difference that is observed is largely due to the values at which the other variables in the model have been set. To reiterate a point made previously, marginal effects are calculated at the point of means while the discrete change estimations are based on a hypothetical male who is in Grade 11 in 2002, matric in 2003 and has no missing information.

It was noted in Section 5.1.2 that the effect of income and ability on educational attainment can only be compared on the basis of a standard deviation change as the variables have different units of measurement. The discussion in this section will focus on this comparison, as well as a brief analysis of the unconditional enrolment estimates.

## Matriculation and conditional enrolment

Comparing the results of Table 5.2.2 to those of Table 5.1.7, both Africans and Coloureds at the matriculation stage are more sensitive to a change in ability than income. Note, however, that the gap between the estimates is greater for Africans than Coloureds. In other words, Africans are noticeably more affected by ability than income while Coloureds only slightly more so.

Interestingly, the pattern is reversed at the enrolment stage for all regressions but one. Africans are slightly more sensitive to income than ability in the first two regressions, but are less sensitive when all variables have been controlled for. As it is assumed that Regression 3 captures current income, it seems that manifest ability is more important than current income at both stages of educational attainment for Africans. Coloureds, on the other hand, are consistently and substantially more responsive to income than ability at the enrolment stage.

**Table 5.2.2: Discrete change in the probability of matric and conditional enrolment for a standard deviation change in LNE score around the mean**

	Matric		PSE   Matric	
	Africans	Coloureds	Africans	Coloureds
<b>Regression 1</b>	0.102	0.217	0.095	0.152
<b>Regression 2</b>	0.091	0.187	0.079	0.106
<b>Regression 3</b>	0.084	0.176	0.077	0.078

Note: The levels of the probabilities from which the changes are calculated are reported in Tables A.6 and A.7 in the Appendix.

The finding that ability is relatively more important at the matriculation stage for both races is not altogether surprising. As discussed already, the direct cost of attending secondary school may not be a major concern due to school fees which are reduced or waived altogether. Due to poor employment opportunities, the opportunity cost for Africans is also very low. As a result, income has little bearing while ability is relatively important. For Coloureds, who do face more substantial opportunity costs, income is only slightly less important than ability in determining matric.

Once a YA has passed Grade 12, the role of ability declines for both races. It is possible that after matric is obtained, ability may play more of a role in determining *where* to enrol as opposed to whether or not to enrol at all. Lam et al's (2010) finding that the LNE score has a greater impact when only universities or technikons are considered, as opposed to all post-secondary institutions, lends support to this suggestion.

The fact that the impact of a standard deviation increase in income on enrolment is almost twice that of ability for Coloureds reflects the relatively high opportunity cost of further education the Coloured group faces. In contrast, ability remains the stronger determinant of educational attainment for Africans at the tertiary level.

### **Enrolment – conditional versus unconditional estimates**

A comparison of the conditional versus unconditional estimates over ability reveals a slightly more complicated story than was found for income. There it was discovered that the estimates using Bayes' Law represented a compromise between the conditional estimates, which may suffer from selection bias, and the single regression unconditional estimates, which ignore the sequential nature of educational attainment in South Africa. Here, the pattern is not as consistent.

With regards to Coloureds, the Bayes' Law results suggest that the role of ability in enrolment decisions is underestimated by both of the other methods. For Africans, however, the conditional estimates consistently underestimate the role of ability while the single regression estimates overestimate the effect in Regressions 1 and 2 and underestimate it in Regression 3. While these results are slightly puzzling, it should be noted that none of the conditional estimates are drastically different to those generated by Bayes' law. As such, sample selection bias is unlikely to be a major issue and the conditional estimates can be used without too much concern.



**Table 5.2.3: Discrete change in the probability of enrolment for a standard deviation change in the LNE score around the mean**

	PSE   Matric		PSE: Single Regression		PSE: Bayes' Law	
	Africans	Coloureds	Africans	Coloureds	Africans	Coloureds
<b>Regression 1</b>	0.095	0.152	0.121	0.133	0.115	0.173
<b>Regression 2</b>	0.079	0.106	0.119	0.071	0.102	0.101
<b>Regression 3</b>	0.077	0.078	0.069	0.051	0.092	0.079

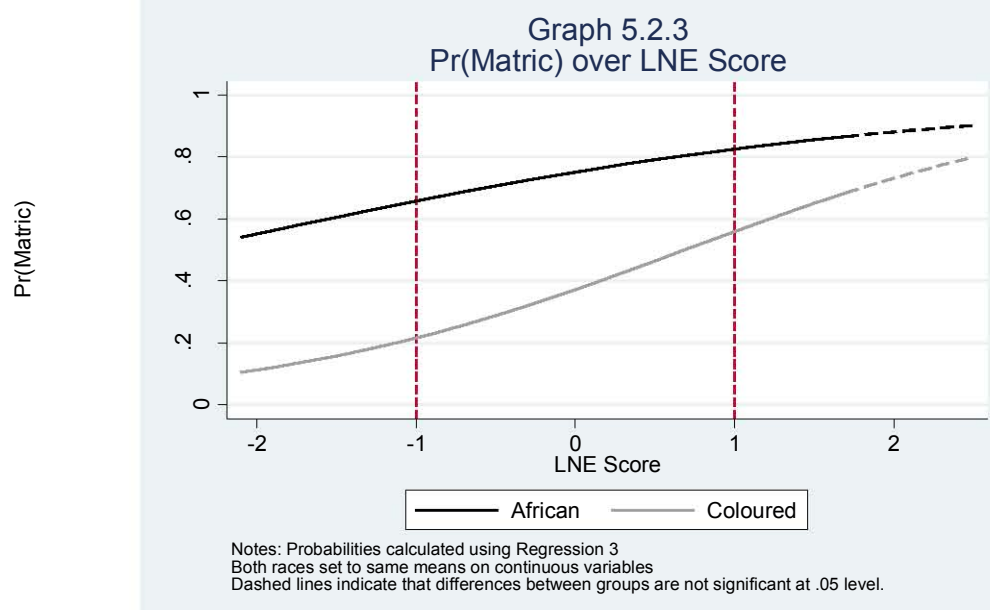
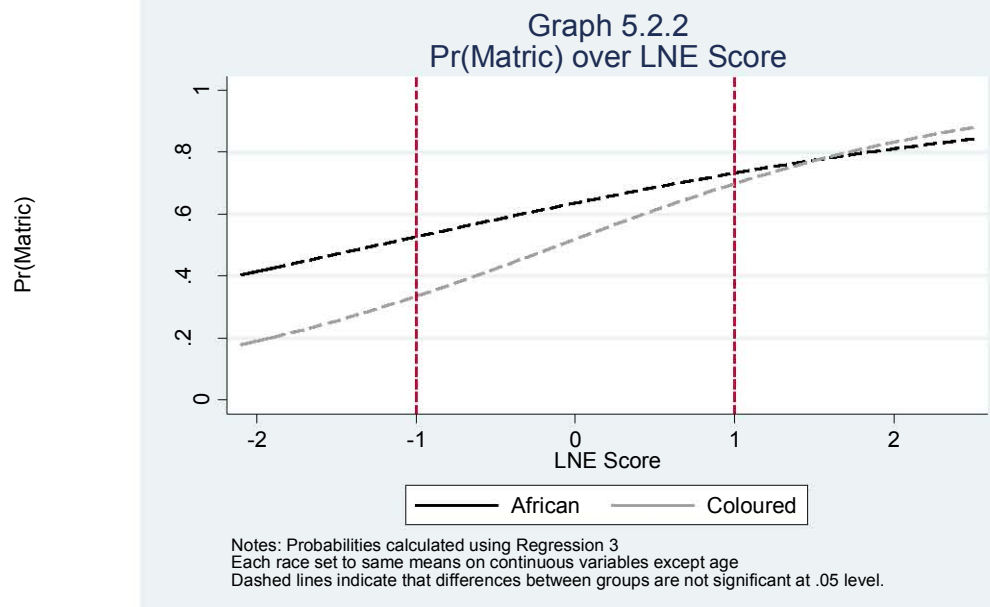
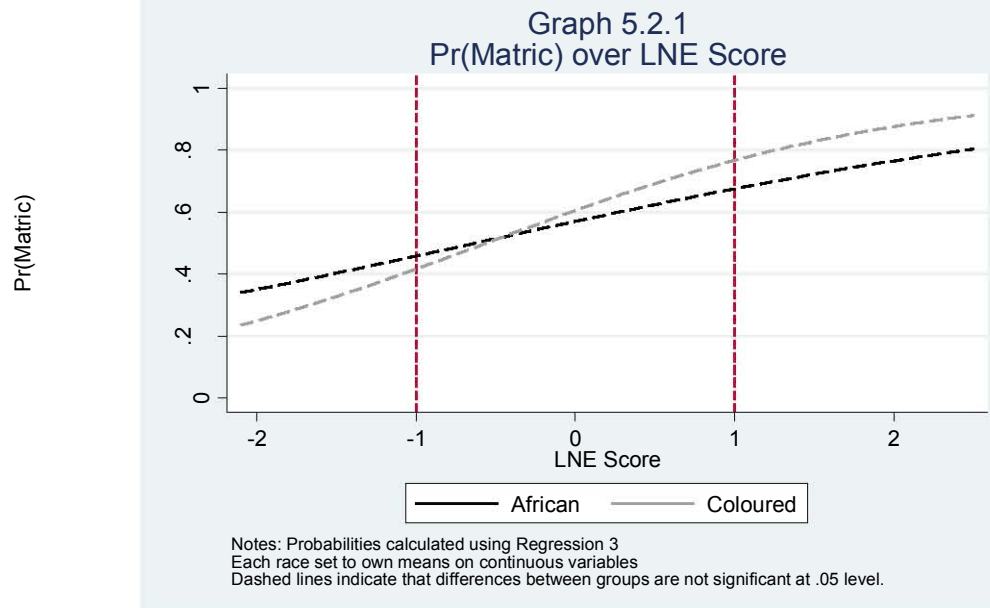
### 5.2.3 Predicted probability graphs

As with the analysis of income, graphs of the predicted probability of matriculation and enrolment over the range of ability are presented below. Once again, it is the racial gap in predicted educational attainment that is of interest, and whether this gap is affected by equating the races on all bases.

#### Stage 1 – Matric

The graphs for matriculation over the range of ability are remarkably similar to those for income. The only substantial difference relates to the slope of the curves, with those for ability being steeper in order to reflect greater sensitivity to ability than income.

Overall, these graphs indicate that Africans and Coloureds have statistically equivalent probabilities of matriculating when their ability is set equal, as well as when all characteristics except age are equalised. Comparing the races when they are held identical reveals that Africans are substantially and significantly more likely to matriculate over the entire range of ability.



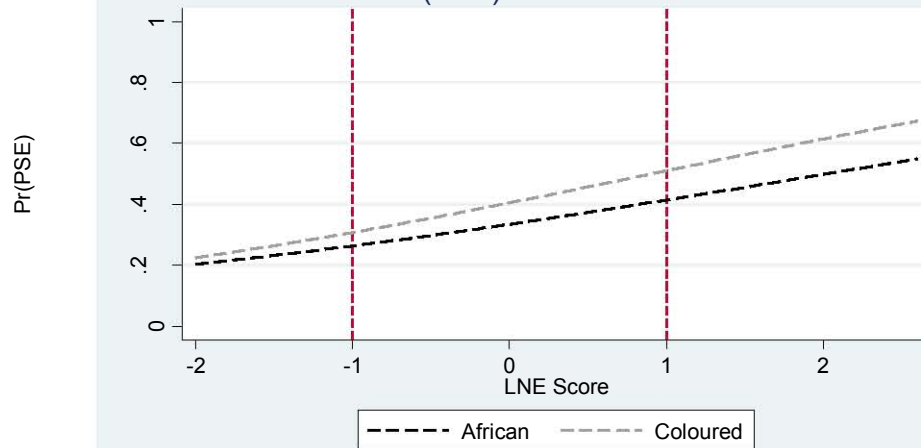
## **Stage 2 - Enrolment into post-secondary education**

In contrast to income, the graphs over ability at the enrolment stage appear quite different to those at matric. This is due to the slopes of the curves being, by and large, parallel across races. This makes sense in light of Regression 3 of Table 5.2.2, where it was revealed that a standard deviation change in ability has an almost identical effect on the probability of enrolment across races.

When the races are held to their own characteristics, Coloureds have an insignificantly higher enrolment rate than Africans. Setting the races equal on all characteristics except age reverses the ranking, and equalising age increases the gap between Africans and Coloureds. However, the difference between the races remains insignificant. Thus while the probability of Africans enrolling at low income values is higher than that of identical Coloureds, at no level of ability does one race exhibit a significant advantage.

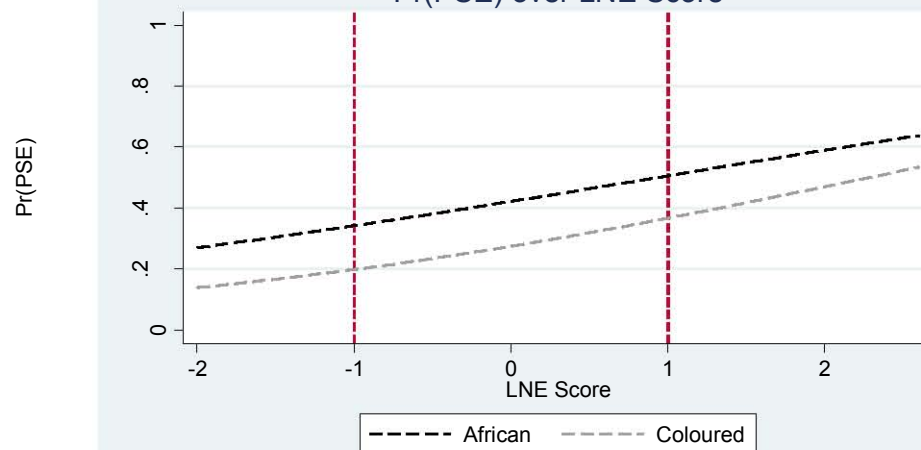
University of Cape Town

Graph 5.2.4  
Pr(PSE) over LNE Score



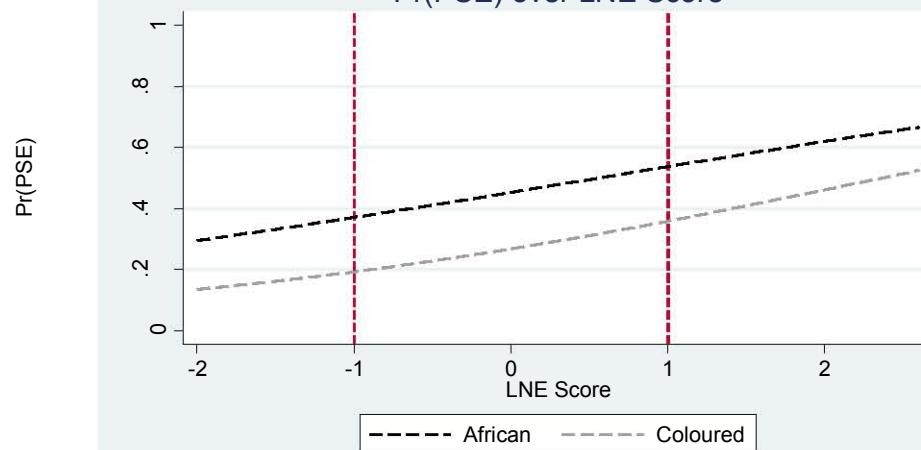
Notes: Probabilities calculated using Regression 3  
Each race set to own means on continuous variables  
Dashed lines indicate that differences between groups are not significant at .05 level.

Graph 5.2.5  
Pr(PSE) over LNE Score



Notes: Probabilities calculated using Regression 3  
Both races set to same means on continuous variables except age  
Dashed lines indicate that differences between groups are not significant at .05 level.

Graph 5.2.6  
Pr(PSE) over LNE Score



Notes: Probabilities calculated using Regression 3  
Each race set to same means on continuous variables  
Dashed lines indicate that differences between groups are not significant at .05 level.

### 5.3 Interactions between income and ability

The final section of this analysis looks at interaction effects between income and ability. It is possible that the role of income in educational attainment varies by the level of ability. For example, if a YA has a low level of ability and believes that further education is out of his reach, current income may be irrelevant in the decision to enrol; the YA will not enrol no matter how high his household income is. Conversely, a high ability individual may choose to enrol even if he is relatively resource constrained due to a strong desire or expectation to do so. Such an individual may also be in a better position to obtain funding through scholarships and bursaries.

As Africans and Coloureds differ significantly in terms of both average income and manifest ability, considering the interaction of these variables may shed more light of their respective processes of educational attainment. In order to test for interaction effects, the probability of matric and enrolment are plotted over the income range for three different levels of ability. The LNE score is set to -1, 0 and 1, representing low, medium and high ability individuals.

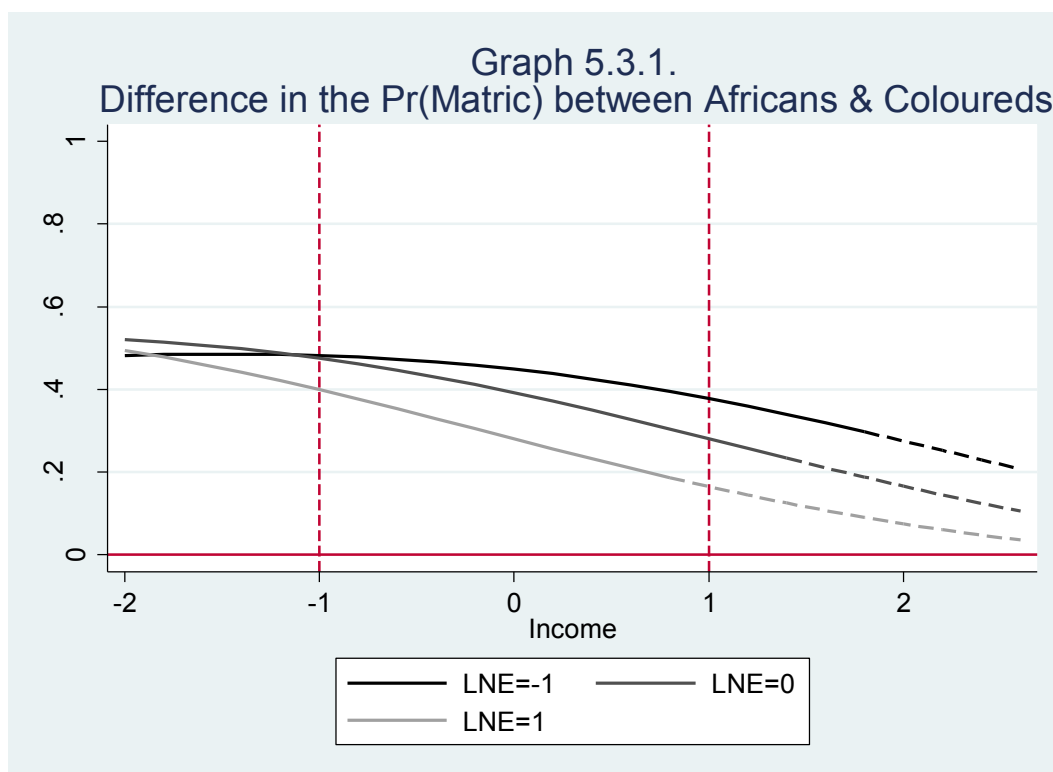
The full graphs showing the predicted probability for each race over the income range are included in the Appendix. The graphs included below summarise the findings onto one set of axes for each stage. Note that the curves represent the *difference* in probability of matriculation/enrolment between races, rather than the levels. Where the line is dashed, the difference is insignificant at the 5 percent level. Thus the black curve in the following graph represents the probability of an African YA matriculating minus the probability of an identical Coloured YA matriculating, when both races are held to an LNE score of -1.

#### Stage 1 - Matric

It is apparent in Graph 5.3.1 that the gap between the races is affected by two factors. Firstly, the difference between the races declines as income increases; this is illustrated by the negative slope of each curve. Secondly, the gap becomes smaller as the level of ability rises, shown by the fact that the curve shifts downwards as the LNE score is increased. While Africans are almost always predicted to have a significantly greater probability of matriculation than Coloureds within the income range [-1;1], the gap is clearly smaller as individuals become wealthier and more able.

This suggests that the process of decision-making occurring within each of the races results in more variation in matriculation rates at lower levels of income and ability. At low income, the relative opportunity cost of staying in school is high for Coloureds, as they stand a good chance of finding

work should they leave school. If, in addition, they have low ability, school work is likely to be more demanding and the possibility of failure is greater. In combination, these factors may encourage a

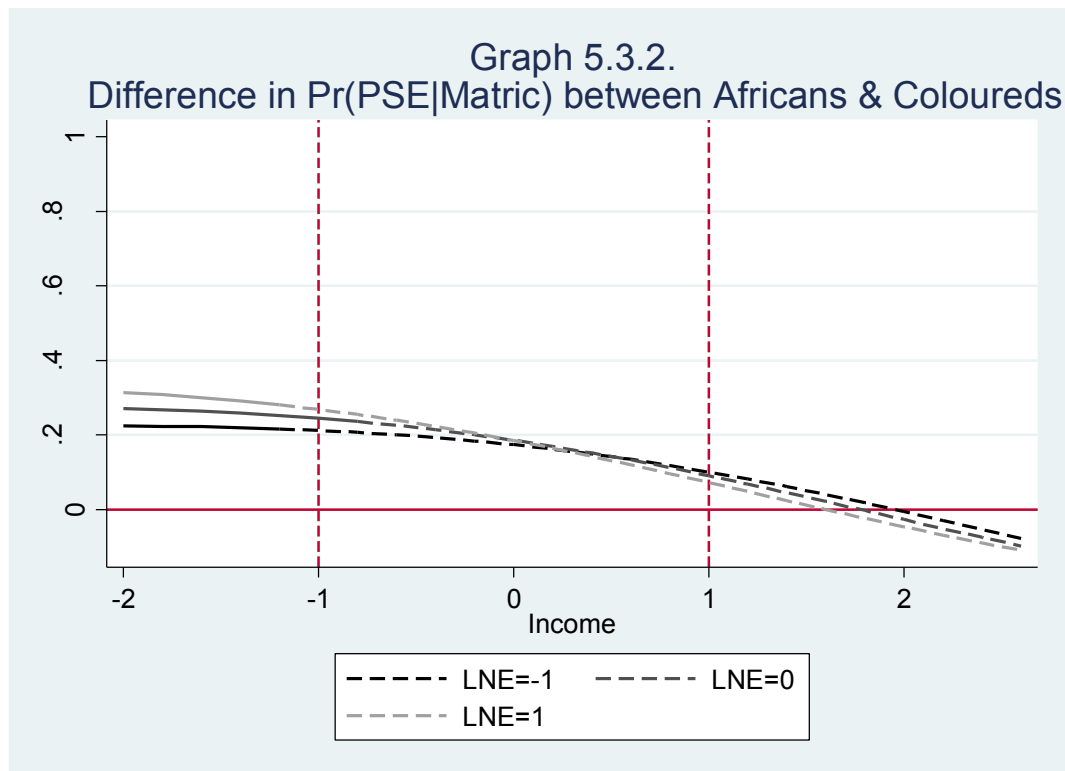


Coloured youth to drop out. Africans with the same characteristics have less incentive to drop out because of dismal labour market prospects. Thus they choose to stay on even if the chance of passing Grade 12 is low.

In contrast, the matriculation rates of high income, high ability Africans and Coloureds are much closer, becoming statistically equivalent at values of 1 on both variables. At this level of ability, school work may not be too demanding and the chance of passing matric is high. Furthermore, the higher household per capita income makes labour market options relatively less attractive for Coloureds. As a result, YAs at the upper end of the income and ability distributions from both races may choose to stay on and complete Grade 12 because there is no need or incentive to drop out. Furthermore, it is plausible that high income, high ability individuals are more aware of the employment and wage returns to further education, and feeling that this might be a possibility in the future, they choose to complete high school. At least with regards to Coloureds, individuals with lower income and ability may believe that PSE is not a viable option for them and therefore choose to enter the labour market as early as possible. Africans of low ability may also believe that PSE is out of reach, but for lack of a better alternative, they remain enrolled in high school.

## Stage 2 – Enrolment into post-secondary education

The final graph looks at the interaction between income and ability at the enrolment stage. In contrast to the matriculation graph, it appears that the gap in enrolment rates is entirely unaffected by the level of ability. The gap does narrow as income rises, but is estimated to be insignificant over the entire range of income within the common support area.



Bearing in mind that the insignificance of these estimates could be due to small sample sizes, the graph indicates that enrolment rates are unlikely to differ significantly between Africans and Coloureds at any point in the ability and income distributions<sup>19</sup>. Thus it seems that the races have enrolment rates which are to all intents and purposes equal whether the races are held to the same characteristics or not. This may not be for the same reason, however. Coloureds are likely to be affected quite strongly by the opportunity cost of further time out of the labour market. Africans, on the other hand, may realise that labour market opportunities with just a matric are limited, but may not believe them to be any better with further education due to the relatively disadvantaged background which they grew up in (Beattie, 2002).

<sup>19</sup> The only exception to this is at the very lower end of the income distribution among individuals with average ability.

## 6. CONCLUSION

This analysis has sought to understand the process of educational attainment among Africans and Coloureds in the Western Cape, focusing on the role of current income and manifest ability. As it is clear that Coloureds are better off in terms of household income, parental characteristics, school quality and employment, it is puzzling that the gap in matriculation rates is not larger and that conditional enrolment rates are statistically equivalent.

When reviewing the literature, it became apparent that Coloureds face markedly better opportunities in the labour market than Africans at levels of education below matric; they find employment more quickly and command a higher wage. As the opportunity cost of time in school is the foregone wage, Coloureds face a substantially higher opportunity cost to remaining in secondary education than Africans. This is likely to explain the finding of this analysis that both races are affected by manifest ability at the matriculation stage, but only Coloureds are significantly affected by current income. Once either income or ability is equalised across the races, the gap in the matriculation rate disappears, and equalising all characteristics results in Africans being substantially and significantly more likely to matriculate than Coloureds.

Turning to the post-secondary stage, among those with matric, the Coloured advantage in terms of employment and wages persists. As a result, it is expected that the labour market continues to have a greater pull on Coloureds, discouraging enrolment in PSE. In contrast, by acquiring a tertiary qualification, Africans face a dramatically increased likelihood of employment and their expected wage triples, both of which are incentives to enrol. As such, the finding that current income again plays a significant role in educational attainment for Coloureds but not Africans makes sense: Coloureds face a non-trivial opportunity cost to enrolling while Africans face a low opportunity cost in combination with a high return.

If these were the only relevant factors in the decision to pursue further studies, one might expect the predicted rate of enrolment of Africans to be significantly greater than that of Coloureds when the races are equalised on all included socio-economic measures. Instead, it is found that the races remain equally likely to enrol, and that this rate remains low. It must be acknowledged that the probabilities at the enrolment stage may be imprecisely estimated due to a relatively small number of observations. However, if the results are taken at face value, they indicate that the differential opportunity cost and payoff to PSE does not lead to a racial gap in enrolment rates. This may be explained by Beattie's suggestion that individuals who grow up in areas with high rates of poverty and unemployment may be partially or fully unaware of the labour market returns to higher



education (2002). This can be directly applied to Africans in the Western Cape, who traditionally live in townships with high rates of poverty and unemployment and are located far away from areas of business. In this case, the ability to afford PSE may never become an issue. Rather, YAs in this scenario are discouraged from applying at the outset, believing that further studies represent an expense and effort which do not bear fruit in the labour market. Thus the role of current income may be insignificant for Africans at this stage not because individuals are financially unconstrained, but rather because enrolling is ruled out before affordability becomes a consideration.

It is beyond the scope of this paper to account for labour market factors or expectations when modeling the process of educational attainment. As such, these factors are put forward merely as suggestions which may explain the findings of this paper. It would be useful if future research could explicitly include these factors in order to highlight the exact role they play in the process of educational attainment by race.

The results indicate that government policies which affect the affordability of education may be ineffective at both the secondary and tertiary level for Africans. At the secondary level, Africans have a relatively high enrolment rate and low dropout rate, implying that it is not attendance that is an issue, but rather the successful completion of high school. In addition, school fees are low and can be waived entirely. Rather, policies which target long-term factors such as household resources and quality of education are more likely to improve the African matriculation rate. This may be the case at the tertiary level as well. In addition, policies which inform Africans who live in poorer areas of the benefits of further education may also promote enrolment.

Given that Coloureds are found to be significantly affected by current income, they may be responsive to a different type of policy. Their relatively low matriculation rate is largely attributable to a high dropout rate, which is assumed to be driven by a desire to enter the labour force. Making education more easily affordable may effectively raise current household income and thereby reduce the relative appeal of the labour market. The same argument can be extended to PSE, as income also has a significant effect on educational attainment at this level.

Finally, as ability is found to play a significant role at both stages for both races, it seems reasonable that any policy which positively influences the development of cognitive and non-cognitive skills, either via early childhood interventions or through the school system, will improve educational attainment for both Africans and Coloureds.

## 7. REFERENCES

- Albert, C. (2000). Higher education demand in Spain: The influence of labour market signals and family background. *Higher Education* 40 , 147-162.
- Alexander, K. L., & Entwisle D R, H. C. (1997). From first grade forward: early foundations of high school dropout. *Sociology of Education* 70(2) , 87-107.
- Anderson, K. G. (2000). *Family structure, parental investment and educational outcomes among Black South Africans*. University of Michigan: PSC Research Report 00-461.
- Anderson, K. G., Case, A., & Lam, D. (2001). Causes and consequences of schooling outcomes in South Africa: Evidence from survey data. *Social Dynamics* 27(1) , 37-59.
- Astone, N. M., & McLanahan, S. S. (1991). Family structure, parental practices and high school completion. *American Sociological Review* 56(3) , 309-320.
- Baumann, R. (2007). *Differences in the college enrollment decision across race* . College of the Holy Cross Faculty Research Series, Paper No. 07-05.
- Beattie, I. R. (2002). Are all "adolescent econometricians" created equal? Race, class and gender differences in college enrolment. *Sociology of Education* 75(1) , 19-43.
- Belley, P., & Lochner, L. (2007, October). *The changing role of family income and ability in determining educational achievement*. Retrieved January 19, 2011, from <http://ssrn.com/abstract=1024131>
- Beutel, A. M., & Anderson, K. G. (2008). Race and the educational expectations of parents and children: The case of South Africa. *The Sociological Quarterly* 49 , 335-361.
- Bhorat, H., & Leibbrandt, M. (2001). Modelling vulnerability and low earnings in the South African labour market. In H. Bhorat, M. Leibbrandt, M. Muziye, S. van den Berg, & I. Woolard, *Fighting poverty: Labour markets and Inequality in South Africa* (pp. 107-129). Cape Town: UCT Press.
- Bhorat, H., & Oosthuizen, M. (2008). Determinants of Grade 12 pass rates in the post-apartheid South African schooling system. *Journal of African Economies* 18(4) , 634-666.
- Branson, N., Leibbrandt, M., & Zuze, L. T. (2009). *The demand for tertiary education in South Africa*. SALDRU Report.
- Bratti, M. (2002). Parents' current income, long-term family characteristics and children's educational outcomes: evidence from the 1970 British Cohort Study. *University of Warwick* .
- Bray, R., Gooskens, I., Kahn, L., Moses, S., & Seekings, J. (2010). *Growing up in the new South Africa: Childhood and adolescence in post-apartheid South Africa*. Cape Town: HSRC Press.
- Cameron, S., & Heckman, J. J. (2001). The dynamics of educational attainment for black, hispanic and white males. *The Journal of Political Economy* 109(3) , 455-499.

Carneiro, P., & Heckman, J. J. (2002). The evidence of credit constraints in post-secondary schooling. *The Economic Journal* 112 , 705-734.

Crouch, L., & Mabogoane, T. (2001). No magic bullets, just tracer bullets: The role of learning resources, social advantage and education management in improving the performance of South African schools. *Social Dynamics* , 60-78.

Crouch, L., & Vinjevold, P. (2006). *South Africa: Access before quality, and what to do now?* Revista de curriculum y formacion del profesorado.

de Broucker, P. (2005). *Getting there and staying there: Low-income students and post-secondary education*. Research Report W|27, Work Network.

Dearden, L., McGranahan, L., & Sianesi, B. (2004). The role of credit constraints in educational choices: evidence from the NCDS and BCS70. *Centre for the Economics of Education, LSE* .

Dias, R. (2005). Education and economic status in South Africa: Insights from the Labour Force Survey of 2003. *Paper presented to the Economics Society of South Africa Biennial Conference*. September 2005.

Downey, D. B. (1995). When bigger is not better: family size, parental resources and children's educational performance. *American Sociological Review* 60(5) , 746-761.

Dustmann, C. (2004). Parental background, secondary school track choice, and wages. *Oxford Economic Papers* 56(2) , 209-230.

Fehrman, P. G., Keith, T. Z., & Reimers, T. M. (1987). Home influence on school learner: Direct and indirect effects of parental involvement on high school grades. *The Journal of Educational Research* 80(6) , 330-337.

Fitzgerald, J., Gottschalk, P., & Moffitt, R. (1998). An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics. *The Journal of Human Resources* 33(2) , 251-299.

Gurgand, M., Lorenceau, A., & Melonio, T. (2011). Student loans: liquidity constraint and higher education in South Africa. *Unpublished manuscript* .

Gustman, A. L., & Steinmeier, T. L. (1981). The impact of wages and unemployment on youth enrollment and labor supply. *The Review of Economics and Statistics* 63(4) , 553-560.

Hansen, K. T., Heckman, J. J., & Mullen, K. J. (2004). The effect of schooling and ability on achievement test scores. *Journal of Econometrics* 121 , 39-98.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labour market outcomes and social behaviour. *AEA annual meeting*. Boston.

Hofmeyr, C. E. (2008). *How has the return to higher education changed in post-apartheid South Africa?* UKZN: Unpublished Honours dissertation.

Hung, F. S., Chung, Y. P., & Ho, E. S. (2000). To work or to continue to higher education? The choice of senior secondary students in Shenzhen, China. *Higher Education* 39(4) , 455-467.

- Jacoby, H. G. (1994). Borrowing constraints and progress through school: Evidence from Peru. *The Review of Economics and Statistics* 76(1) , 151-160.
- Kodde, D. A. (1986). Uncertainty and the demand for education. *The Review of Economics and Statistics* 68(3) , 460-467.
- Kodde, D. (1988). Unemployment expectations and human capital formation. *European Economic Review* 32(8) , 1645-1660.
- Kodde, D., & Ritzen, J. M. (1988). Direct and indirect effects of the parental education level on the demand for higher education. *The Journal of Human Resources* 23(3) , 356-371.
- Lam, D. (1999). *Generating extreme inequality: Schooling, earnings and intergenerational transmission of human capital in South Africa and Brazil. Research Report*. Population Studies Centre at the Institute for Social Research, University of Michigan.
- Lam, D., & Leibbrandt, M. (May 10, 2009). Schooling inequality and earnings inequality in South Africa since the end of apartheid. *Prepared for 2009 RC28 Conference, Beijing Conference* .
- Lam, D., Ardington, C., & Leibbrandt, M. (2008). *Schooling as a lottery: racial differences in school advancement in urban South Africa*. University of Michigan: Population Studies Centre research Report 08-632.
- Lam, D., Ardington, C., Branson, N., Goostrey, K., & Leibbrandt, M. (April, 2010). Credit constraints and the racial gap in post-secondary education in South Africa. *Population Association of America*. Dallas, Texas.
- Lam, D., Leibbrandt, M., & Mlatsheni, C. (2007). Education and youth unemployment in South Africa. *IPC Working Paper Series Number 34* .
- Lam, D., Leibbrandt, M., & Mlatsheni, C. (2009). Human capital, job search and unemployment among young people in South Africa. *Preliminary draft for the IZA/World Bank on Employment and Development, May 2010*. Cape Town.
- Lauer, C. (2003). Family background, cohort and education: A French–German comparison based on a multivariate ordered probit model of educational attainment. *Labour Economics* 10(2) , 231-251.
- Long, J. S. (2009). Group comparisons in logit and probit using predicted probabilities. *Unpublished manuscript* .
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks: Sage Publications, Inc.
- Long, J. S., & Freese, J. (2006). *Regression models for categorical dependent variables using Stata*. College Station: Stata Press.
- Moll, P. G. (1998). Primary schooling, cognitive skills and wages in South Africa. *Economica* 65(258) , 263-284.
- Mora, J. G. (1997). Equity in Spanish higher education. *Higher Education* 33 , 233-249.

Tomkowicz, J., & Bushnik, T. (2003). Who goes to post-secondary education and when: pathways chosen by 20 year-olds. *Statistics Canada* .

van der Berg, S. (2006). Public spending and the poor since the transition to democracy. In H. Borat, & R. Kanbur, *Poverty and Policy in Post-Apartheid South Africa*. Pretoria: HSRC Press.

Willis, R. J., & Rosen, S. (1979). Education and self-selection. *The Journal of Political Economy* 87(5) , S7-S36.

Xu, J., & Long, J. S. (2005). Confidence intervals for predicted outcomes in regression models for categorical outcomes. *The Stata Journal* 5 , 537-559.

Yamauchi, F. (2005). Race, equity and public schools in post-apartheid South Africa: Equal opportunity for all kids. *Economics of Education Review* 24 , 213-233.

University of Cape Town

## 8. APPENDIX

**Table A.1: Discrete change in the probability of matriculation for a standard deviation change in income around the mean**

	Africans			Coloureds		
	Pr(Matric) <sub>A</sub>	Pr(Matric) <sub>B</sub>	ΔPr(Matric)	Pr(Matric) <sub>A</sub>	Pr(Matric) <sub>B</sub>	ΔPr(Matric)
<b>Regression 1</b>	0.687	0.751	0.063	0.367	0.572	0.205
<b>Regression 2</b>	0.718	0.770	0.052	0.313	0.487	0.174
<b>Regression 3</b>	0.744	0.781	0.037	0.320	0.463	0.142

Notes:

**Table A.2: Discrete change in the probability of enrolment given matriculation for a standard deviation change in income around the mean**

	Africans			Coloureds		
	Pr(PSE) <sub>A</sub>	Pr(PSE) <sub>B</sub>	ΔPr(PSE)	Pr(PSE) <sub>A</sub>	Pr(PSE) <sub>B</sub>	ΔPr(PSE)
<b>Regression 1</b>	0.489	0.605	0.116	0.322	0.583	0.261
<b>Regression 2</b>	0.493	0.595	0.102	0.273	0.504	0.231
<b>Regression 3</b>	0.428	0.464	0.037	0.249	0.381	0.132

Notes:

**Table A.3: Discrete change in the unconditional probability of enrolment for a standard deviation change in income around the mean, using the full sample**

	Africans			Coloureds		
	Pr(PSE) <sub>A</sub>	Pr(PSE) <sub>B</sub>	ΔPr(PSE)	Pr(PSE) <sub>A</sub>	Pr(PSE) <sub>B</sub>	ΔPr(PSE)
<b>Regression 1</b>	0.255	0.384	0.129	0.073	0.225	0.152
<b>Regression 2</b>	0.279	0.394	0.115	0.056	0.162	0.106
<b>Regression 3</b>	0.253	0.319	0.066	0.056	0.113	0.058

Notes:

**Table A.4: Discrete change in the unconditional probability of enrolment for a standard deviation change in income around the mean, using Bayes' Law**

	Africans			Coloureds		
	Pr(PSE) <sub>A</sub>	Pr(PSE) <sub>B</sub>	ΔPr(PSE)	Pr(PSE) <sub>A</sub>	Pr(PSE) <sub>B</sub>	ΔPr(PSE)
<b>Regression 1</b>	0.336	0.454	0.118	0.097	0.299	0.202
<b>Regression 2</b>	0.353	0.458	0.105	0.070	0.218	0.148
<b>Regression 3</b>	0.318	0.363	0.045	0.071	0.161	0.091

Notes:

**Table A.5 : Means used in predicted probability graphs**

	Each race set to own means		Both races set to same means
	Africans	Coloureds	Africans and Coloured
Age in 2005: Matriculation	21.57	19.97	20.77
: Enrolment	20.84	19.91	20.38
LNE score	-0.18	0.51	0.19
Mother's education	7.46	8.68	8.12
Father's education	4.58	6.89	5.83
Prop. life lived with mother	0.79	0.91	0.86
Prop. life lived with father	0.44	0.67	0.57
Number of siblings	2.56	1.80	2.15

**Table A.6: Discrete change in the probability of matriculation for a standard deviation change in the LNE score around the mean**

	Africans			Coloureds		
	Pr(Matric) <sub>A</sub>	Pr(Matric) <sub>B</sub>	ΔPr(Matric)	Pr(Matric) <sub>A</sub>	Pr(Matric) <sub>B</sub>	ΔPr(Matric)
Regression 1	0.646	0.748	0.102	0.348	0.565	0.217
Regression 2	0.685	0.777	0.091	0.293	0.480	0.187
Regression 3	0.708	0.791	0.084	0.288	0.465	0.176

Notes:

**Table A.7: Discrete change in the probability of enrolment given matriculation for a standard deviation change in the LNE score around the mean**

	Africans			Coloureds		
	Pr(PSE) <sub>A</sub>	Pr(PSE) <sub>B</sub>	ΔPr(PSE)	Pr(PSE) <sub>A</sub>	Pr(PSE) <sub>B</sub>	ΔPr(PSE)
Regression 1	0.449	0.545	0.095	0.401	0.552	0.152
Regression 2	0.509	0.588	0.079	0.271	0.377	0.106
Regression 3	0.419	0.496	0.077	0.243	0.322	0.078

Notes:

**Table A.8: Discrete change in the unconditional probability of enrolment for a standard deviation change in the LNE score around the mean, using the full sample**

	Africans			Coloureds		
	Pr(PSE) <sub>A</sub>	Pr(PSE) <sub>B</sub>	ΔPr(PSE)	Pr(PSE) <sub>A</sub>	Pr(PSE) <sub>B</sub>	ΔPr(PSE)
Regression 1	0.210	0.331	0.121	0.111	0.245	0.133
Regression 2	0.268	0.387	0.119	0.066	0.137	0.071
Regression 3	0.224	0.293	0.069	0.055	0.106	0.051

Notes:

**Table A.9: Discrete change in the unconditional probability of enrolment for a standard deviation change in the LNE score around the mean, using Bayes' Law**

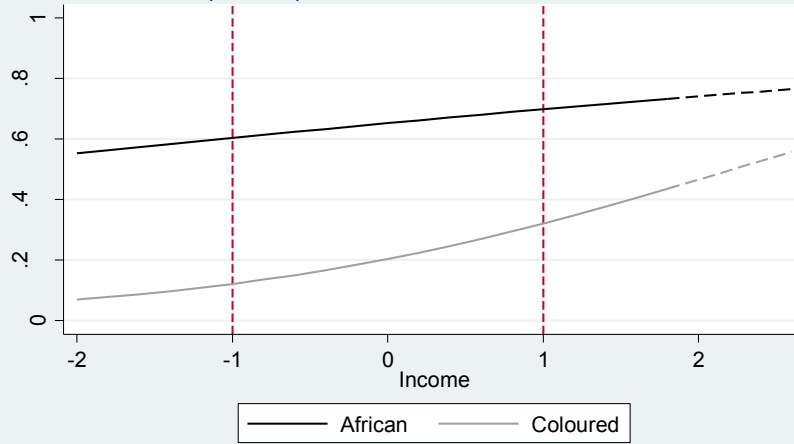
	Africans			Coloureds		
	$\Pr(\text{PSE})_A$	$\Pr(\text{PSE})_B$	$\Delta\Pr(\text{PSE})$	$\Pr(\text{PSE})_A$	$\Pr(\text{PSE})_B$	$\Delta\Pr(\text{PSE})$
<b>Regression 1</b>	0.339	0.454	0.115	0.140	0.312	0.173
<b>Regression 2</b>	0.400	0.502	0.102	0.079	0.181	0.101
<b>Regression 3</b>	0.337	0.429	0.092	0.070	0.149	0.079

Notes:

University of Cape Town

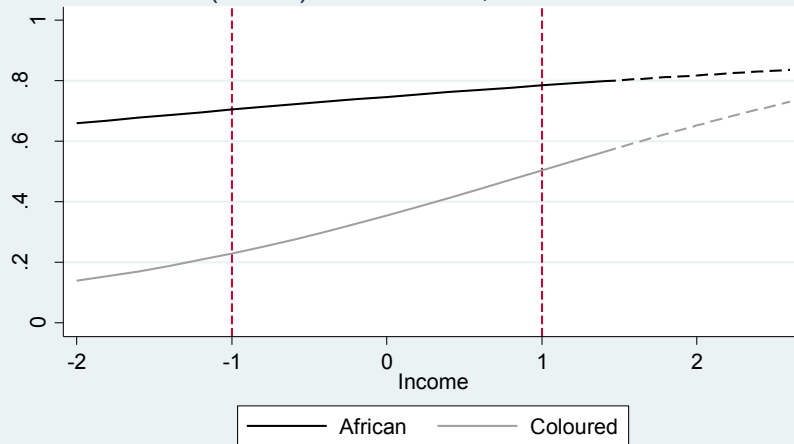


Graph A.1  
Pr(Matric) over Income, LNE Score = -1



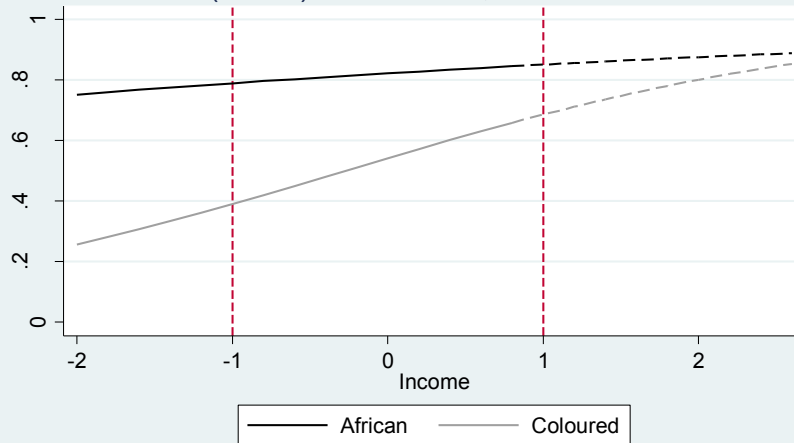
Notes: Probabilities calculated using Regression 3  
Races set to same means on continuous variables  
Dashed lines indicate that differences between groups are not significant at .05 level.

Graph A.2  
Pr(Matric) over Income, LNE Score = 0



Notes: Probabilities calculated using Regression 3  
Races set to same means on continuous variables  
Dashed lines indicate that differences between groups are not significant at .05 level.

Graph A.3  
Pr(Matric) over Income, LNE Score = 1



Notes: Probabilities calculated using Regression 3  
Races set to same means on continuous variables  
Dashed lines indicate that differences between groups are not significant at .05 level.

